



Transmissão de Vídeo em Tempo Real na Internet

**Rui Jorge dos Santos
Almeida**



Transmissão de Vídeo em Tempo Real na Internet

**Rui Jorge dos Santos
Almeida**

Dissertação apresentada à Universidade de Aveiro para cumprimento dos requisitos necessários à obtenção do grau de Mestre em Engenharia Electrónica e Telecomunicações, realizada sob a orientação científica do Dr. Armando José Formoso de Pinho, Professor Associado do Departamento de Electrónica e Telecomunicações da Universidade de Aveiro

o júri

presidente

Doutor José Luís Guimarães Oliveira
Professor Associado da Universidade de Aveiro

Doutor Armando José Formoso de Pinho
Professor Associado da Universidade de Aveiro (orientador)

Doutor Pedro Manuel Quintas Aguiar
Professor Auxiliar do Instituto Superior Técnico da Universidade Técnica de Lisboa

agradecimentos

Desejo agradecer ao Prof. Dr. Armando Pinho a orientação, a preciosa ajuda e a permanente disponibilidade prestada ao longo de todo o trabalho.

Ao meu colega de Mestrado, José Antunes, pela colaboração que tornou possível a conclusão do primeiro ano do mestrado.

Ao colega António José Neves, no IEETA, pela imprescindível e abnegada ajuda durante a fase experimental.

Este trabalho é dedicado à minha esposa e ao meu filho, cuja paciência e compreensão ao longo de todo o tempo em que me dediquei à elaboração desta dissertação foi o maior incentivo à persecução dos objectivos que me foram propostos.

resumo

O presente trabalho aborda o tema da transmissão de vídeo em tempo real na Internet. Ao longo do mesmo, são enunciadas e explicadas as principais dificuldades colocadas pela Internet para a transmissão deste tipo de multimédia. As principais evoluções tecnológicas que pretendem atenuar estas dificuldades e permitir a implementação destes sistemas são identificadas, principalmente no que diz respeito à evolução das normas de codificação e compressão como, por exemplo, as normas MPEG, e à implementação da tecnologia de *Streaming* de vídeo. As novas tecnologias que ainda estão em desenvolvimento mas que prometem melhorar o desempenho dos sistemas de transmissão de vídeo em tempo real são também abordadas. Incluímos um capítulo com experiências práticas reais de transmissão de vídeo em tempo real na Internet com diferentes condições de codificação e de transmissão e respectivos resultados.

A transmissão de vídeo em tempo real pode vir a ser a principal aplicação para entretenimento ou fins profissionais da Internet a médio prazo. Para isso, muito têm concorrido os vários desenvolvimentos tecnológicos, nomeadamente ao nível da codificação e compressão de vídeo. Em poucos anos, devido à contribuição de normas como o MPEG, o H.263, o AVC ou aos codificadores proprietários, o tamanho final dos ficheiros de vídeo codificados diminuiu mais de 75%. Estão em fase de implantação novas normas, como o MPEG-7 e MPEG-21, que facilitarão o trabalho de pesquisa e de adaptação às condições individuais de cada utilizador e às suas próprias preferências. Se conjugarmos este factor com o crescimento das velocidades de acesso à Internet, por exemplo ao nível da Banda Larga com o xDSL ou por Cabo, e com o crescimento da capacidade dos computadores pessoais ou dos telemóveis, apenas será necessário que a vertente comercial esteja adequada às capacidades dos consumidores para que a transmissão de vídeo na Internet com qualidade seja, não só possível, como também uma realidade para a generalidade das pessoas.

abstract

This work intends to present a global view about the transmission of real-time video over the Internet. We present and explain the main problems caused by the architecture of the actual Internet. The principal technical solutions that seek to overcome these problems are introduced, with the focus on the evolution of compression developments like, among others, the MPEG standards, as well as on the explanation of Streaming video technology. We also address some of the new standards that might help to solve and improve transmission of this media. We include a practical chapter with the results of tests for real-time video transmission over the Internet for different conditions of compression and transmission.

Transmission of real-time video over the Internet may become, in a few years, the main application for entertainment or business purposes. This objective has received strong contribution from technical evolutions, mainly in video compression. With the development of standards like MPEG, H.263, AVC or with the proprietary codecs, the length of the encoded video bitstream has been reduced over 75% in a short period of time. There are new ways that other tools, like MPEG-7 or MPEG-21, have been exploiting to find new forms of helping the user to find and get the best result according to his technical conditions and personal preferences. Additionally, there has been an important growing on the number of Broadband accesses, like xDSL or Cable modem, and on the capacities of PC's and other devices that supports Streaming video software. The last piece in this picture is the price of the Internet access and the software. If they are affordable to the consumer's earnings, there is no doubt that quality real-time video over the Internet is possible and that it will be a reality.

Índice:

CAPÍTULO I – A INTERNET E A TRANSMISSÃO DE VÍDEO EM TEMPO REAL	11
I.1 – A INTERNET	13
I.2 - PEQUENA HISTÓRIA DA INTERNET	13
I.3 – DESCRIÇÃO FUNCIONAL DA REDE.....	14
I.4 – FORMAS DE ACESSO À REDE.....	16
I.5 – PRINCIPAIS CONSTRANGIMENTOS À TRANSMISSÃO DE VÍDEO EM TEMPO REAL.....	18
CAPÍTULO II - CODIFICAÇÃO DE VÍDEO	21
II.1 CODIFICADORES CLÁSSICOS.....	22
II.1.1 <i>Compensação de Movimento</i>	23
II.1.2 <i>Transformada Discreta de Co-seno (DCT)</i>	27
II.1.3 <i>Quantificação e Codificação por VLC</i>	28
II.1.4 - <i>Organização do bitstream</i>	29
II.2 PRIMEIRAS NORMAS DE CODIFICAÇÃO DE VÍDEO: H.261, H.263 E MPEG-1 E MPEG-2	30
II.2.1 <i>H.261</i>	30
II.2.2 <i>H.263</i>	31
II.2.3 <i>MPEG-1 e MPEG-2</i>	35
II.3 - MPEG-4	39
II.3.1 - <i>Modo de Funcionamento</i>	40
II.3.2 <i>Principais funcionalidades para adaptação à Internet</i>	42
II.4 ADVANCED VIDEO CODING (H.264 OU MPEG-4 PART 10).....	46
II.4.1 <i>Codificação (VCL)</i>	47
II.4.2 <i>Adaptação ao meio de transmissão (NAL)</i>	49
II.5 – CODECS PROPRIETÁRIOS.....	49
II.5.1 <i>QuickTime</i>	49
II.5.2 <i>RealNetworks</i>	50
II.5.3 <i>Windows Media</i>	51
CAPÍTULO III - STREAMING DE VÍDEO	55
III.1 - FUNCIONAMENTO DO STREAMING	56
III.1.1 <i>Codificação de Vídeo adaptada a streaming</i>	57
III.1.2 <i>Codecs de Vídeo para Streaming de Média</i>	59
III.1.3 <i>Servidores para Streaming</i>	59
III.1.4 <i>Escalabilidade e Adaptação da taxa de transmissão</i>	61
III.1.5 <i>Mecanismos de distribuição de streaming multimédia</i>	61
III.1.6 <i>Sincronização de média</i>	62
III.2 DESENVOLVIMENTOS NA COMPRESSÃO DE VÍDEO.....	63
III.2.1 <i>Escalabilidades do tipo SNR, espacial e temporal</i>	63
III.2.2 <i>Fine Granularity Scalability (FGS)</i>	64
III.2.3 <i>Progressive Fine Granularity Scalability (PFGS)</i>	66
III.2.4 <i>Técnicas de compressão híbridas Temporal/FGS</i>	67
III.3 DESENVOLVIMENTOS NO CONTROLO DE QUALIDADE DE SERVIÇO.....	68
III.3.1 <i>Controlo de congestionamento</i>	68
III.3.2 <i>Controlo de erros</i>	68
III.4 DESENVOLVIMENTO DE SERVIÇOS DE DISTRIBUIÇÃO CONTÍNUOS	69
III.4.1 <i>Filtragem pela rede</i>	69
III.4.2 <i>Multicast do nível de aplicação</i>	69
III.4.3 <i>Replicação de conteúdos</i>	71
III.4.4 <i>Redes Distribuídas</i>	71
III.5 DESENVOLVIMENTO DE PROTOCOLOS ADAPTADOS PARA STREAMING NA INTERNET.....	72
III.5.1 <i>Protocolo de rede</i>	72
III.5.2 <i>Protocolos de transporte</i>	73

III.5.3 Protocolos de sessão.....	73
III.6 CONCLUSÃO.....	74
CAPÍTULO IV – O FUTURO DO VÍDEO EM TEMPO REAL NA INTERNET	75
IV.1 - NOVOS SERVIÇOS NA INTERNET.....	76
<i>Serviços Integrados - IntServ</i>	<i>76</i>
<i>Serviços diferenciados - Diffserv</i>	<i>76</i>
IV.2 - DESENVOLVIMENTOS NOS ELEMENTOS DE REDE E FORMAS DE ACESSO	77
IV.3 - NOVAS TÉCNICAS PARA STREAMING DE VÍDEO NA INTERNET	78
<i>Eficiência e Utilização do AVC na Internet.....</i>	<i>78</i>
IV.4 – ACESSO MULTIMEDIA UNIVERSAL, ESCALABILIDADE E TRANSCODIFICAÇÃO, MPEG-7 E MPEG-21	79
<i>Escalabilidade e Transcodificação.....</i>	<i>80</i>
MPEG-7	82
MPEG-21	84
CAPÍTULO V – EXPERIÊNCIAS PRÁTICAS DE TRANSMISSÃO DE VÍDEO NA INTERNET	87
V.1 – HARDWARE E FERRAMENTAS DE SOFTWARE	87
V.2 – DESCRIÇÃO DAS EXPERIÊNCIAS.....	88
V.3– RESULTADOS	90
a) <i>Windows Media Encoder, leitor no PC A ligado por ADSL.....</i>	<i>90</i>
b) <i>Windows Media Encoder, leitor no PC A ligado por Modem.....</i>	<i>91</i>
c) <i>Windows Media Encoder, leitor no PC B ligado por ADSL.....</i>	<i>92</i>
d) <i>Windows Media Encoder, leitor no PC B ligado por Modem.....</i>	<i>93</i>
e) <i>Real Producer, leitor no PC A ligado por ADSL.....</i>	<i>94</i>
f) <i>Real Producer, leitor no PC A ligado por Modem.....</i>	<i>95</i>
g) <i>Real Producer, leitor no PC B ligado por ADSL.....</i>	<i>96</i>
h) <i>Real Producer, leitor no PC B ligado por Modem.....</i>	<i>97</i>
V.3 – ANÁLISE DOS RESULTADOS.....	98
<i>Influência da velocidade de acesso.....</i>	<i>98</i>
<i>Influência da capacidade do PC.....</i>	<i>103</i>
<i>Comparação entre codificadores.....</i>	<i>104</i>
CAPÍTULO VI – CONCLUSÕES	111
GLOSSÁRIO	116
BIBLIOGRAFIA.....	117
ÍNDICE REMISSIVO	120

Índice de figuras:

<i>Fig. I-1 – Esquema simplificado da transmissão de vídeo na Internet.....</i>	<i>12</i>
<i>Fig. I-2 – Esquema geral da Internet</i>	<i>15</i>
<i>Fig II-1a – Duas imagens consecutivas de um trecho de vídeo [29].....</i>	<i>23</i>
<i>Fig II-1b – Diferença entre as duas imagens consecutivas exibidas na fig. II-1a [29].....</i>	<i>24</i>
<i>Fig II-1c – Vectores de movimento entre imagens consecutivas [29].....</i>	<i>24</i>
<i>Fig II-1d – Informação residual após o processo de Compensação de movimento [29].....</i>	<i>25</i>
<i>Fig II-2 – Processo de criação do vector de movimento [27].....</i>	<i>26</i>
<i>Fig II-3 - Codificador com Compensação de Movimento.....</i>	<i>27</i>
<i>Fig. II-4 - Exemplo de Construção do bitstream por varrimento em zig-zag.....</i>	<i>28</i>
<i>Fig II-5 – Codificador Clássico com Compensação de Movimento, DCT e Quantificação</i>	<i>29</i>
<i>Fig II-6 - Codificador típico para H.263.....</i>	<i>32</i>
<i>Fig II-7 – Codificador típico para H.263++</i>	<i>34</i>
<i>Fig II-8 - Codificador típico para MPEG-2</i>	<i>37</i>
<i>Fig II-9 - Decomposição da imagem em Video Objects (VO's) [26]</i>	<i>41</i>
<i>Fig II-10 - Escalões para as ferramentas de codificação no MPEG-4</i>	<i>44</i>
<i>Fig II-11 – Codificador típico MPEG-4</i>	<i>45</i>
<i>Fig. II-12 - Exemplo da aplicação de Codificação por Sprites.....</i>	<i>46</i>
<i>Fig II-13 - Codificador típico AVC.....</i>	<i>48</i>
<i>Figura III-1 – Esquema geral de streaming de vídeo.....</i>	<i>56</i>
<i>Fig. III-2 – Qualidade/Velocidade para compressões escaláveis e não escaláveis.....</i>	<i>58</i>
<i>Fig III-3 – Codificador FGS</i>	<i>65</i>
<i>Fig. III-4 – Decodificador FGS</i>	<i>66</i>
<i>Fig III-5 – Distribuição de Vídeo sem Multicast pela Rede.....</i>	<i>70</i>
<i>Fig III-6 – Distribuição de vídeo numa Rede com IP Multicast.....</i>	<i>70</i>
<i>Fig. III-7 - protocolos numa sessão de streaming.....</i>	<i>72</i>
<i>Fig. IV-1 – Esquema geral do conceito Universal Multimedia Access.....</i>	<i>80</i>
<i>Fig. IV-2 – Transcodificação para adaptação a diferentes condições de transmissão ou para necessidade de alteração do tipo de codificação.....</i>	<i>82</i>
<i>Fig. IV-3 – Esquema geral para um sistema com MPEG-7</i>	<i>84</i>
<i>Fig. IV-4 – Sistematização dos componentes da norma MPEG-21</i>	<i>86</i>
<i>Fig. V-1 – Esquema da experiência prática.....</i>	<i>87</i>
<i>Fig. V-2 – Exemplo de uma frame do vídeo original á entrada do codificador.....</i>	<i>98</i>
<i>Fig. V-3 – Exemplo de frame codificada pelo Windows Media Encoder a 40 Kbps e recebida via modem....</i>	<i>98</i>
<i>Fig. V-4 – Exemplo de frame codificada pelo Real Producer Basic a 50 Kbps e recebida via modem.....</i>	<i>99</i>
<i>Fig. V-5 – Exemplo de frame distorcida. Codificação pelo Windows Media Encoder a 40 Kbps e recebida via modem.....</i>	<i>99</i>
<i>Fig. V-6 – Exemplo de frame distorcida. Codificação pelo Real Producer Basic a 50 Kbps e recebida via modem.....</i>	<i>100</i>
<i>Fig. V-7 – Exemplo de frame do vídeo original á entrada do codificador.....</i>	<i>100</i>
<i>Fig. V-8 – Exemplo de frame codificada pelo Windows Media Encoder a 491 Kbps e recebida via ADSL. 101</i>	
<i>Fig. V-9 – Exemplo de frame codificada pelo Real Producer Basic a 450 Kbps e recebida via ADSL.....</i>	<i>101</i>
<i>Fig. V-10 – Exemplo de frame com sobreposição de imagens. Codificação pelo Windows Media Encoder a 691 Kbps e recebida via ADSL.....</i>	<i>102</i>
<i>Fig. V-11 – Exemplo de frame distorcida. Codificação pelo Real Producer Basic a 700 Kbps e recebida via ADSL</i>	<i>102</i>
<i>Fig. V-12 – Comparação da percentagem de frames recebidas por cada PC ligado por ADSL</i>	<i>103</i>
<i>Fig. V-13 –Comparação da percentagem de frames recebidas por cada PC ligado por modem.....</i>	<i>104</i>
<i>Fig. V-14 – Exemplo de frame do vídeo original á entrada do codificador.....</i>	<i>106</i>
<i>Fig. V-15 – Exemplo de imagem obtida por codificação a 491 Kbps pelo Windows Media Encoder.....</i>	<i>106</i>
<i>Fig. V-16 – Exemplo de imagem obtida por codificação a 450 Kbps pelo Real Producer Basic.....</i>	<i>107</i>
<i>Fig. V-17 – Exemplo de uma frame do vídeo original á entrada do codificador.....</i>	<i>108</i>
<i>Fig. V-18 – Exemplo de imagem obtida por codificação a 491 Kbps pelo Windows Media Encoder.....</i>	<i>108</i>
<i>Fig. V-19 – Exemplo de imagem obtida por codificação a 450 Kbps pelo Real Producer Basic.....</i>	<i>109</i>
<i>Fig VI-1 - Evolução dos débitos gerados pelas normas de codificação de vídeo [3].....</i>	<i>112</i>

Capítulo I – A Internet e a Transmissão de Vídeo em Tempo Real

O crescimento desmesurado da Internet e a sua consolidação como um meio eficaz de comunicação global para toda a humanidade, deixou de ser um objectivo utópico para se tornar numa realidade incontestável, permitindo-lhe ser capaz de colocar em contacto pessoas ou entidades em qualquer parte do mundo. Hoje, a Internet deixou definitivamente de estar confinada aos meios universitários ou institucionais e, através dos serviços que suporta, está progressivamente a ocupar um lugar privilegiado na nossa vida privada e nos meios profissionais. Este meio de comunicação suporta aplicações e serviços que já são indispensáveis para a actividade empresarial ou para a vida diária particular tais como o correio electrónico, vulgo *e-mail* ou a consulta de informação em páginas *web*. Um inquérito realizado por uma empresa norte-americana de software a alguns executivos e responsáveis informáticos de médias e grandes empresas [5], revelou que a perda do e-mail pode ser mais traumatizante do que um divórcio...

A Internet teve origem numa aplicação de comunicação privada militar norte-americana em meados dos anos 70, com o objectivo de interligar informaticamente algumas instalações dispersas geograficamente. O conceito foi mais tarde transportado para uma escala global e veio a garantir a partilha de informação entre todos os pontos do mundo. As formas de aceder a esta rede mundial são bastante diversas, o que tem a vantagem de facilitar e generalizar o acesso, mas também o inconveniente de criar uma rede heterogénea em termos de resposta e comportamento.

Devido à sua vasta implementação geográfica, a Internet foi vista quase de imediato como um meio de distribuição de eventos em tempo real na forma de transmissão em directo de imagem e som. As aplicações são tão diversas como, por exemplo, uma videoconferência, o fornecimento de serviços de *video-on-demand*, o ensino à distância, as bibliotecas digitais ou simplesmente o entretenimento. O desenvolvimento de aplicações específicas para a apresentação de vídeo em *PC's* (Computadores Pessoais) ou mesmo em *PDA's* (*Personal Digital Assistant*), permitiu difundir largamente a tecnologia e tornou os utilizadores finais receptivos a estas aplicações, abrindo também a necessidade de corresponder às suas expectativas cada vez mais exigentes. Hoje, é perfeitamente vulgar a transmissão de eventos desportivos, concertos musicais ou palestras pela Internet com milhares de clientes em simultâneo. Por outro lado, a criação de canais comerciais de TV na Internet está em franco crescimento e contribuirá, num curto espaço de tempo, para o aumento da procura deste tipo de tecnologia.

Numa visão muito global, poderíamos descrever o processo de transmissão de um vídeo na Internet da seguinte forma: as imagens do evento a transmitir são captadas e digitalizadas por câmaras adequadas. Os ficheiros de vídeo resultantes têm a particularidade de serem enormes o que dificulta a sua transmissão ao longo de uma rede em que não há uma taxa de transmissão garantida, constituída por receptores completamente distintos ao nível da ligação e desempenho de processamento dos dados como é o caso da Internet. Seria absolutamente impossível a sua transmissão em bruto e um dos grandes desafios que os investigadores desta área têm de suplantar é o da necessidade de compressão destes ficheiros. Os ficheiros, ou são enviados imediatamente pela rede ao cliente ou ficam armazenados em servidores que estão acessíveis via Internet para que, após um pedido do receptor, sejam enviados ao longo da rede. O receptor tem uma aplicação desenvolvida

para a leitura destes ficheiros cuja função é a de descomprimi-los e apresentá-los em imagens (ver Fig I-1).

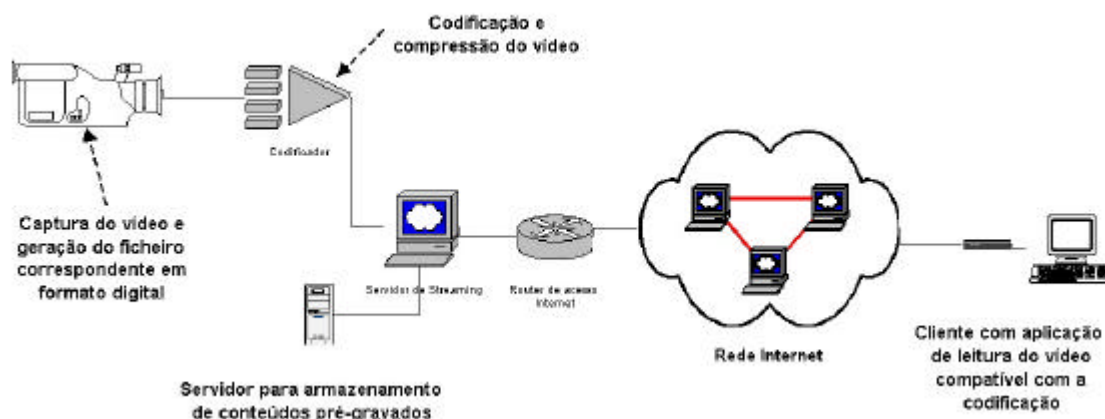


Fig. I-1 – Esquema simplificado da transmissão de vídeo na Internet

No entanto, esta visão simplista esbarra na arquitectura da Internet e nos meios de acesso postos à disposição dos utilizadores finais. Mais especificamente, as características do sinal de vídeo exigem do canal de transmissão um conjunto de requisitos bastante exigentes tais como necessidade de largura de banda e tempos de atraso muito pequenos, para garantir a continuidade da apresentação ao receptor humano que não deseja ver interrupções ou paragens no visionamento, e pequenas taxas de perda de pacotes para uma reconstrução eficaz da imagem final. No entanto, a Internet actual, sendo uma rede de *best effort*, não é capaz de o garantir, quer devido às características dos componentes da rede, tais como os equipamentos activos ou as ligações, quer devido às questões protocolares subjacentes à própria rede.

O objectivo deste trabalho é apresentar os vários constrangimentos associados a esta tecnologia e analisar as soluções que presentemente estão a ser estudadas, de forma a poder apontar caminhos de evolução para as mesmas.

Para colmatar ou reduzir alguns dos problemas existentes, vários esforços têm vindo a ser feitos para desenvolver soluções que tornem fiável e aplicável a transmissão de vídeo em tempo real na Internet. O *Streaming* de vídeo, tecnologia que será desenvolvida neste trabalho, tem características que permitem superar algumas destas dificuldades, nomeadamente a do tempo necessário para a apresentação do objecto, pois permite ao utilizador final não ter de esperar pela chegada completa do ficheiro, ou as perdas ao longo da rede, dado possuir alguns mecanismos de detecção de erros e controlo de fluxo que minimizam ou ultrapassam as eventuais perdas de pacotes e atrasos na rede. Esta tecnologia ainda está em fase de desenvolvimento, não estando ainda suficientemente madura para se considerar absolutamente fiável.

1.1 – A Internet

A Internet é uma ferramenta indissociável do mundo actual. Esta rede global de comunicação foi capaz de criar e fortalecer laços de contacto entre indivíduos ou entre empresas, independentemente do espaço físico ou cultural que os separe. A transmissão de vídeo em tempo real nesta rede surgiu quase naturalmente por necessidade dos próprios utilizadores que pretendiam utilizar as características de universalidade para terem acesso a eventos em qualquer parte do mundo. Nos próximos pontos deste Capítulo pretende-se fazer uma descrição do actual estado da rede mundial e identificar os seus pontos fortes e fracos no que diz respeito à implementação de aplicações de vídeo em tempo real.

1.2 - Pequena história da Internet

Na década de 60 do século passado, em plena guerra fria, os meios militares norte-americanos debatiam-se com a necessidade de partilharem informação em tempo real, de forma segura, sustentada por uma rede capaz de resistir a um ataque nuclear, dado que a principal limitação da tradicional rede telefónica era que a queda de um circuito cortava as comunicações dos clientes sustentados nesse troço [1]. Foi então criada a ARPA (*Advanced Research Projects Agency*) que, através de colaborações com universidades e algumas empresas, se dedicou à construção de uma rede de comutação de pacotes em que cada terminal estava ligado a pelo menos outras 2 máquinas da rede para que, no caso da queda de um troço, as comunicações fossem automaticamente encaminhadas por um percurso alternativo. Originalmente, utilizaram-se circuitos a 56 Kbps alugados às companhias telefónicas para interligar os computadores numa arquitectura de servidores e clientes, com protocolos desenvolvidos especificamente para garantir as comunicações entre todos os elementos da rede. Assim surgiu a ARPANET que começou a crescer através da progressiva integração de instituições de ensino superior norte-americanas que por sua vez contribuíam com investigação e desenvolvimento para a melhoria do desempenho da mesma, incluindo a possibilidade de ligação via rádio ou satélite.

Quando o crescimento obrigou à implementação de subredes, foi necessário um protocolo específico para a interligação das mesmas e surgiu o modelo TCP/IP (*Transport Control Protocol / Internet Protocol*), que foi considerado o seu protocolo oficial em 1 de Janeiro de 1983, e posteriormente de um sistema de correspondência entre nomes e endereços, o DNS (*Domain Name System*).

Dado o enorme sucesso e fiabilidade desta rede, a National Science Foundation (NSF) nos EUA teve a ideia de criar uma rede semelhante para interligar todas as instituições de ensino, desta vez sem a necessidade de existência de acordos com a ARPA, para a troca de informações científicas e prevendo a ligação de investigadores a nível individual instalados em suas casas ou a museus, laboratórios e bibliotecas. Surgiu então a NSFNET, que manteve a ligação à ARPANET mas, por questões de segurança, sem a componente de ligação às instituições militares. O crescimento desta rede foi exponencial, e pela primeira vez, foi utilizado um *backbone* em fibra óptica com PC's IBM a servirem de *Routers*.

O passo seguinte foi a entrada de empresas que exploraram comercialmente a NSFNET que acabou por ser vendida à *America On Line* (AOL). Por outro lado, na Europa, esta ideia de criação de uma rede IP foi seguida com a criação da *EuropaNET*, sendo posteriormente criadas redes nacionais em cada país que se interligavam mutuamente.

Desta forma começou-se a falar frequentemente de uma “Internet” até o termo se tornar vulgar na década de 80 para designar a rede IP de dimensão global.

O seu crescimento tem sido desmesurado: em 1990 existiam cerca de 300 000 utilizadores, em 1992 atingiu-se o primeiro milhão. Por cada ano que passou na década de 90 o seu tamanho duplicou e foram sendo diversificadas as formas de ligação com a integração da ligação via redes móveis, por exemplo [1].

A Internet, com todas as suas vantagens e desvantagens, é o veículo de aplicações importantíssimas hoje em dia, tais como a vulgar navegação em Páginas Web, o correio electrónico, vulgo *e-mail*, a transferência de ficheiros através do FTP (*File Transfer Protocol*), a conversa escrita em tempo real, através do IRC (*Internet Relay Chat*) e, mais importante para este trabalho, a transmissão de eventos em vídeo e/ou áudio.

A transmissão de vídeo na Internet é comercialmente muito interessante para um fornecedor de conteúdos de vídeo, pois pode difundir-los numa rede vasta onde encontra numerosos potenciais clientes. Por outro lado, do ponto de vista do cliente, é uma forma de aceder a eventos que ocorrem em sítios distantes e que, de outra forma, estaria impossibilitado de lhes aceder. Entre as várias aplicações possíveis podemos indicar:

- ? Videoconferência;
- ? Video on-demand;
- ? Transmissão de eventos em directo, como por exemplo concertos musicais ou eventos desportivos;
- ? Ensino à distância;
- ? Sistemas de vigilância remota;
- ? Acesso a documentos na forma de bibliotecas multimédia;

1.3 – Descrição funcional da rede

A Internet pode ser definida tecnicamente como uma rede de máquinas (computadores pessoais, servidores, PDA's, telemóveis de terceira geração, etc.) que [2]:

- ? utilizam o protocolo TCP/IP ou UDP
- ? têm um endereço IP
- ? têm a capacidade de enviar e receber pacotes de e para qualquer máquina na Internet

Cada máquina está ligada a um Internet Service Provider (ISP) que o coloca em contacto com o resto da rede. Esta ligação pode ser feita de formas bastante distintas quer ao nível físico (cabo ou *wireless*) quer ao nível da velocidade de ligação (desde alguns Kbps até aos Gbps), como também no que diz respeito ao acordo comercial entre a entidade que se liga e o ISP.

A rede não tem, globalmente, uma estrutura pré-definida e pode ser vista como uma interligação de subredes de circuitos de alto débito (*backbones*), tipicamente em fibra óptica, e *routers* de alta capacidade. A estes backbones ligam-se redes mais pequenas, ao nível regional, com cabos e routers menos potentes e que por sua vez suportam as ligações aos utilizadores finais, quer sejam PC's individuais ou LAN's de empresas ou escritórios (ver Fig I-2).

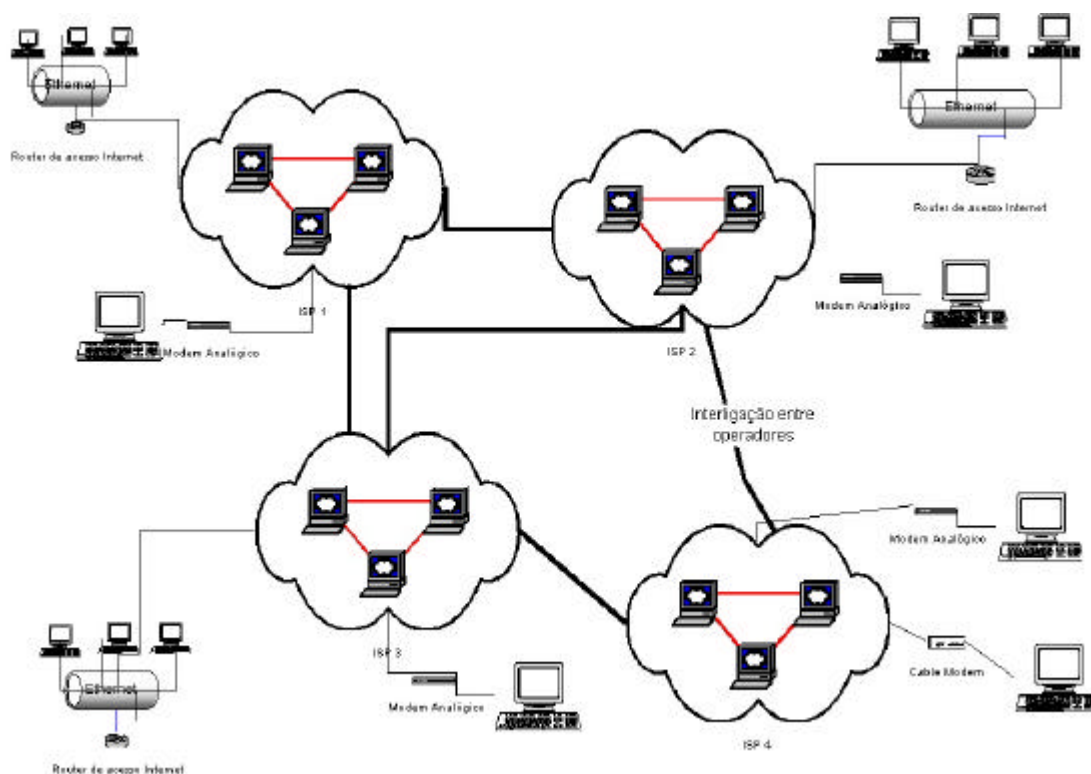


Fig. I-2 – Esquema geral da Internet

O grande responsável pela boa ligação de todos estes elementos é o protocolo **IP (Internet Protocol)**, desenvolvido com o grande objectivo de fazer chegar pacotes de dados entre dois elementos mesmo que estejam fisicamente em redes diferentes.

Cada mensagem enviada é dividida em pequenos pacotes, e cada um destes recebe uma etiqueta que identifica, entre outros, a sua origem, o seu destino e o tipo de dados. A identificação de cada elemento da rede é feita univocamente através do seu endereço IP. Um endereço IP é constituído por 32 bits na versão ainda corrente denominada IPv4 e passará a 48 bits na versão IPv6 que está a ser desenvolvida. O endereço IP permite também obter informação da rede a que pertence a máquina que o detém. Os Routers ao longo da rede podem encaminhar cada um dos pacotes por caminhos distintos, em função do estado da rede, tal como tinha sido condição na génese da ARPANET, embora não seja garantida a entrega de todos os pacotes (em situações de congestionamento alguns podem ser descartados).

O **Transport Control Protocol (TCP)** é um protocolo orientado à ligação e que foi concebido para garantir uma ligação fiável extremo a extremo numa rede constituída por troços completamente distintos e que, por si só, não oferece garantias. Cabe ao TCP fazer com que pacotes perdidos na rede possam ser reenviados, evitar congestionamentos e também reorganizar os pacotes recebidos dado que geralmente não chegam na ordem em que foram enviados.

Outro dos protocolos de transporte é o **User Data Protocol (UDP)**. O UDP é um protocolo não orientado à ligação, mais leve do que o TCP, e que se utiliza em situações em que se pretende privilegiar a velocidade da rede e onde não há grandes necessidades de segurança

e redundância da informação transmitida. Não garante a entrega dos pacotes, mas detecta a sua perda. Por ser mais simples e ocupar menos recursos da rede é o preferido por aplicações de transmissão de vídeo como será apresentado ainda neste Capítulo.

O protocolo **HTTP (Hiper Text Transfer Protocol)**, utilizado na World Wide Web, não é apropriado para *streaming* vídeo em tempo real, pois embora esteja suportado em TCP e isso lhe permita garantir a entrega da informação, não oferece estabilidade na taxa de transmissão. As primeiras soluções para este problema implicam a existência no receptor de um *buffer* denominado *preroll buffer*, cuja função é a de funcionar como uma barragem num rio de forma que a sua saída tenha um fluxo constante. Desta forma, se definirmos um tempo de *buffer* entre os 5 e os 20 segundos, teremos uma entrega de dados de uma forma constante, pois tipicamente não se registam atrasos superiores na transmissão da informação. A implementação do *buffer* é feita na aplicação de leitura do vídeo instalada no receptor.

No capítulo III, iremos apresentar alguns protocolos específicos que se aplicam à transmissão de vídeo na Internet que foram desenvolvidos com o fim de suplantarem as limitações do TCP e do UDP.

1.4 – Formas de acesso à rede

A enorme heterogeneidade da rede é um dos pontos fracos para a distribuição de vídeo. O facto de cada utilizador ter condições quase únicas faz com que os fornecedores de vídeo na Internet tenham de desenvolver soluções que incluam algoritmos que permitam servir clientes com diferentes velocidades de transmissão da informação e, principalmente, que possam adaptar-se a essa velocidade através de variações da qualidade do vídeo final.

De entre a multiplicidade de formas de ligação de um utilizador à rede podemos identificar as principais:

1. **GSM** : Trata-se da possibilidade de aceder à Internet por um terminal móvel suportado na rede GSM. Por exemplo, pode ser feito ligando um PC dotado de uma placa do tipo PCMCIA a um telemóvel. As velocidades alcançadas não ultrapassam os 9.6 Kbps o que torna muito difícil a transmissão de vídeo em tempo real por este tipo de ligação.
2. **GPRS**: O GPRS, conhecida por geração 2.5 das redes móveis, utiliza a capacidade de agregação de canais de transmissão da rede GSM para aumentar as velocidades de ligação, prometendo uma ligação até 171,2 Kbps. No entanto, a velocidade de ligação depende do número de utilizadores de voz e dados no raio de acção da sua célula, o que em termos práticos significa que não há uma velocidade fixa de ligação e, tipicamente em meios urbanos, as velocidades não são famosas. Ainda está em fase embrionária a implementação da rede UMTS pelos operadores móveis que pretende revolucionar a utilização de aplicações multimédia nas redes móveis.
3. **Ligação via linha telefónica analógica com modem** : Os modems analógicos atingem, via ligações *dial*, os 56 Kbps, muito melhor que os 9.6 ou 22 Kbps dos primeiros tempos desta tecnologia. Embora o modem possa suportar essa velocidade, há outros factores que influenciam a verdadeira velocidade da ligação, tais como a qualidade do par de cobre que liga o utilizador à sua central telefónica e as capacidades dos POPs (Point of Presence) do seu ISP. O acesso analógico à rede telefónica está espalhado por quase todos os lares, o que o faz ser ainda o principal

meio de acesso à Internet. Os fornecedores de conteúdos de vídeo que optam por fornecer diferentes versões do vídeo consoante o tipo de ligação ainda costumam utilizar os 56 Kbps de ligação como uma das referências para as versões disponíveis. Genericamente, a qualidade que se obtém não é famosa, mas para um utilizador pouco exigente já permite uma razoável implementação da tecnologia de vídeo em tempo real na Internet.

4. **Ligação via linha telefónica RDIS:** Uma linha RDIS é um acesso Digital que funciona em canais individuais de 64 Kbps com capacidade de agregação dos 2 canais permitindo ligações a 128 Kbps, necessitando de um modem digital que implemente a norma RDIS. Trata-se de uma ligação com qualidade superior à da linha analógica com a consequente melhoria de qualidade das aplicações de vídeo.
5. **ADSL (Assymetric Digital Subscriber Line) :** É uma tecnologia em franca expansão comercial e que permite rentabilizar os pares de cobre utilizados na rede telefónica analógica ou RDIS para fornecer um acesso de alta velocidade assimétrico, ou seja, com diferentes velocidades de *upload* e de *download*. Em Portugal, o operador de telecomunicações Portugal Telecom fornece aos Internet Service Providers os seguintes serviços, denominados classes de serviço:

Classe	Upload	Download
0	128 Kbps	512 Kbps
1	128 Kbps	256 Kbps
2	128 Kbps	768 Kbps
3	256 Kbps	1024 Kbps

Tabela I – classes de acesso ADSL

Convém referir que as velocidades acima indicadas são as máximas e que, mais uma vez, não existe velocidade garantida de transmissão em todo o troço até à entrada na Internet, dado que as condições variam em função da quantidade de clientes ligados em simultâneo.

O número de utilizadores de serviços de vídeo na Internet com ADSL tem vindo a aumentar pois esta ligação oferece velocidades interessantes com uma componente comercial acessível.

6. **Cabo:** Os operadores de redes de transmissão de televisão por cabo passaram a rentabilizar a capacidade instalada nas suas redes através do fornecimento de ligação à Internet. Em termos de velocidade de acesso as mesmas variam desde os 128 Kbps até 1 Mbps, embora sujeitas a flutuações em função da utilização da rede. Tal como o ADSL, fornece a possibilidade de acesso a conteúdos multimédia com velocidades razoáveis e um custo interessante.
7. **Acesso dedicado:** Entende-se por acesso dedicado à Internet um canal que liga directamente um PC ou uma LAN a um ponto de entrada na Internet e que se caracteriza por ser exclusivamente dedicado a essa ligação com velocidades constantes e garantidas ao longo deste percurso. Pode ser feito por circuito dedicado, em que o operador de Telecomunicações instala uma linha directa entre o cliente e a porta de acesso à Internet, por *Frame Relay*, em que o acesso à Internet é feita via o acesso a uma rede IP de transmissão de pacotes que por sua vez tem conectividade à Internet, em ATM (*Asynchronous Transfer Mode*) ou outro e têm uma gama de oferta de velocidades que variam desde os simples 64 Kbps até 155 Mbps ou mais. Ao contrário dos serviços anteriores, todo o percurso entre o

utilizador e a Internet é exclusivo desta ligação pois não há partilha com outros utilizadores, nem a infra-estrutura de acesso ao cliente é dividida para comunicações de voz ou de televisão, por exemplo. Assim, pode-se definir uma velocidade constante até à porta Internet que lhe foi dedicada. Tipicamente, é este o tipo de ligação à Internet de instituições ou empresas que pretendem ter um meio fiável, seguro e com garantias de largura de banda para a ligação das suas LAN's à rede e/ou para fornecimentos de serviços a essa rede, tais como armazenamento de páginas WWW, domiciliação de servidores de *e-mail* ou FTP, etc..

Para além do tipo de ligação, temos de levar em conta os aspectos comerciais da utilização do serviço e que podem influenciar as decisões dos utilizadores dos serviços de vídeo. Há serviços cuja tarifação depende do tempo de ligação à rede, como as ligações via modem analógico ou RDIS, outras que taxam a quantidade de informação recebida independentemente do tempo da ligação, como o ADSL ou o Cabo, e outras que são do tipo *flat-rate*, ou seja, uma taxa fixa independente da informação recebida ou do tempo de ligação. Assim, o pagamento do serviço pode influenciar a qualidade do vídeo que um determinado cliente pretende, pois este pode preferir ver um vídeo de qualidade inferior para reduzir os seus custos.

1.5 – Principais constrangimentos à transmissão de vídeo em tempo real

Tendo em conta os aspectos apresentados nos pontos anteriores, podemos identificar os seguintes principais problemas da transmissão de vídeo em tempo real na Internet [2]:

A Internet não foi desenhada para conteúdos de média

A Internet é uma rede de transmissão de pacotes suportada pelo protocolo IP. A informação a enviar tem de ser fraccionada em pequenos pacotes que são transmitidos pela rede por um percurso desconhecido à partida e que pode ser diferente para 2 pacotes do mesmo ficheiro. Por outro lado, não há garantia nem da entrega dos pacotes nem da manutenção da ordem aquando da chegada. Assim, a transmissão de eventos contínuos tem de se adaptar a esta realidade, criando mecanismos próprios de redundância ou recuperação da informação perdida e algoritmos capazes da reconstrução do ficheiro inicial sem perda de continuidade e manutenção da sequência original.

A Internet não oferece QoS e regista flutuações no seu comportamento

O protocolo IP não permite, só por si, a possibilidade de garantir Qualidade de Serviço para uma determinada aplicação ou utilizador. A Internet é uma rede *Best Effort*, ou seja, os recursos disponíveis são utilizados pelas aplicações que os solicitam sem preferências. Por outras palavras, se tivermos num rede capaz de suportar 128 Kbps e um utilizador, esse poderá utilizar a totalidade dos 128 Kbps, mas quando a rede é partilhada por outros utilizadores, a largura de banda vai sendo distribuída pelos utilizadores consoante os pedidos forem chegando sem ser feita qualquer cedência de privilégios a uma aplicação mais necessitada do que outra em largura de banda. Em situações de congestionamento, o atraso dos pacotes cresce e em casos extremos a rede descarta mesmo alguns pacotes indistintamente. O resultado é uma rede em que as condições de transmissão de informação para um dado utilizador variam em função do tipo de ligação, da partilha do meio com os outros utilizadores e do volume de tráfego cursado em cada momento. Não há hipótese

alguma de uma velocidade constante e pré-definida. Este facto é determinante na transmissão de vídeo pois esta aplicação consome grandes recursos de largura de banda e é muito exigente com os tempos de atraso e perda de pacotes.

A Internet é um meio heterogéneo

Foi visto atrás que o manancial de formas de ligação à Internet contribuiu para o rápido crescimento e globalização da rede. O reverso da medalha é a heterogeneidade das características de ligação dos utilizadores. Cada um destes tem uma ligação que se distingue dos outros utilizadores principalmente pelo débito que consegue atingir. Mesmo em clientes que utilizem a mesma forma de acesso, p.e. cabo, as variações podem ser significativas devido à quantidade de utilizadores em simultâneo na área de residência.

Outra forma de diferenciação é o tipo de máquina que suporta a aplicação de leitura do vídeo. Cada vez mais se alarga o leque de equipamentos com capacidade de ligação à Internet e onde se tentam implementar leitores de *streaming* de vídeo, entre eles:

- ? Computadores pessoais, com processadores que podem ter grandes diferenças de velocidades de processamento ou diferentes capacidades de memória.
- ? PDA's que implementam *streaming* mas muito mais limitados que os PC's.
- ? Telemóveis, que nos dias de hoje já começam a suportar aplicações de vídeo e cujo desenvolvimento aponta para a convergência com os computadores.

Os fornecedores de conteúdos têm de satisfazer os vários clientes, quer através da disponibilização de diferentes versões do mesmo evento ou através de algoritmos de escalabilidade que possam tirar benefício das ligações mais rápidas e fornecer um serviço mínimo aceitável para os que estão em piores condições. Esta segunda hipótese é, sem dúvida, a opção onde mais intensamente estão a ser feitos desenvolvimentos na área da transmissão de vídeo na Internet. O objectivo é produzir uma única versão do objecto com a capacidade para se adaptar às condições de transmissão e recepção.

Resumindo, a Internet é uma rede densamente distribuída com a capacidade de chegar a quase todas as pessoas no mundo e que pode ser um meio de distribuição de vídeo em tempo real. Contudo, como não foi desenvolvida para esse fim, sendo heterogénea e sem oferta de garantias no seu comportamento, coloca obstáculos à sua implementação que os investigadores desta área têm de resolver.

Ao longo deste trabalho, serão identificadas as técnicas nesta área. O segundo capítulo servirá para descrever o processo da codificação e compressão de vídeo para transmissão na Internet. As principais normas de codificação de vídeo e os principais codificadores proprietários comerciais completarão esta parte do trabalho. Será dado relevo ao MPEG 4 e o seu desenvolvimento posterior o *Advanced Video Coding* (AVC). Seguidamente, no Capítulo III, serão abordados os principais desenvolvimentos técnicos pensados para a transmissão de vídeo na Internet. Será apresentado o *Streaming* de vídeo como a forma mais apropriada de o fazer e também algumas adaptações que foram feitas na rede para o servir. Reservámos o quarto Capítulo desta dissertação para mencionar alguns caminhos de futuro que já começam a ser explorados para o aperfeiçoamento da transmissão de multimédia na Internet. No capítulo V descrevemos um conjunto de experiências práticas realizadas para verificar o quanto é fiável e possível a transmissão de vídeo na Internet e os seus resultados. Por fim, no Capítulo VI, será feito um resumo do trabalho e serão apresentadas algumas conclusões sobre esta matéria.

Capítulo II - Codificação de vídeo

O maior problema à transmissão de vídeo na Internet é o tamanho do ficheiro gerado após o registo das imagens e sons do evento que se pretende transmitir. Por exemplo, se não for feito qualquer tipo de compressão, 5 minutos de sinal de televisão com tamanho 720x576 a 25 imagens por segundo no formato 4:2:0, que necessita de 124 Mbps por segundo [28], ocupam a impressionante quantidade de 4,54 Gbytes!

A transmissão deste tipo de dados pela Internet, nas condições actuais da rede, tornaria a sua implementação bastante penosa para os utilizadores. Como exemplo, o tempo de transmissão do trecho atrás referido para um cliente que utilize uma ligação ADSL a 512 Kbps, mesmo que se considerasse que todos os outros troços da rede estariam dedicados a este utilizador, seria de 20 horas!

É por isto que grande parte do esforço de desenvolvimento está centrado no aspecto da codificação de vídeo com o intuito de ultrapassar esta limitação.

Este passo, fundamental para garantir não só um meio de entendimento entre o emissor e o receptor, mas também para contribuir para a eficiência da transmissão, transforma as imagens e o áudio, enormes consumidoras de recursos em bruto, em trechos de informação mais compactos e mais adaptados ao canal Internet. As desvantagens são a perda de qualidade da imagem final, seja por diminuição do seu tamanho, por menor número de imagens por unidade de tempo ou por uma menor qualidade da mesma.

A forma de compressão das imagens para posterior codificação e transmissão via Internet tem de obedecer a algumas exigências quando comparadas com os algoritmos utilizados pela transmissão de televisão, nomeadamente uma grande escalabilidade, leveza para tratamento pelos PC's e capacidade de compensar as inevitáveis perdas de pacotes e atrasos da rede.

Para resolver estas situações, temos de adaptar a transmissão às condições específicas de cada cliente. Cabe a cada um destes informar e instruir o servidor acerca dos parâmetros mais importantes que lhe estão associados.

A codificação também se deve adaptar ao facto de a transmissão ser em directo ou em diferido, dado que as duas têm diferentes características. Enquanto a primeira tem de permitir o acesso de um cliente durante a transmissão, a segunda tem de implementar as funções de *fastforward*, *rewind*, *pause*, etc.

Finalmente temos de levar em conta a possibilidade de criar formas de codificação capazes de se adaptar aos diferentes tipos de processadores que podem receber a informação tais como *PC's*, *PDA's*, *settop boxes* e outros, que suportam diferentes complexidades, obrigando a codificações escaláveis.

Assim, há diferentes modelos de codificação que variam entre os mais simples, em que, por exemplo, a escalabilidade se resume à criação de diferentes versões cada uma optimizada para um canal específico de transmissão, até às mais complexas em que a codificação se pode chamar de escalável e é baseada em representações espaço-temporais do sinal fonte ou na diferenciação dos degraus de quantificação.

Neste Capítulo, iremos falar da evolução das técnicas de codificação de vídeo, desde o H.261 até às mais recentes evoluções das normas MPEG ou dos Codecs proprietários como o QuickTime, o Windows Media ou o Real Player.

II.1 Codificadores Clássicos

A codificação de vídeo é feita a partir da análise temporal e/ou espacial das imagens primitivas. O seu objectivo é tirar proveito de algumas características do sinal vídeo que permitem subtrair informação redundante, pouco importante ou mesmo tirar partido do funcionamento do sistema de visão humana [1].

Podemos resumir estas propriedades da seguinte forma:

- ? **Redundâncias espaciais ou temporais:** a grande maioria dos pixels de uma imagem tem um valor que não é independente dos pixels vizinhos dentro da mesma imagem ou do correspondente pixel das imagens anteriores e posteriores. Por exemplo, é possível criar sistemas de predição entre imagens consecutivas por forma a que seja apenas necessário enviar para cada uma a diferença em relação à anterior, reduzindo assim a quantidade de informação a ser transportada na rede.
- ? **Redundância estatística:** A utilização de codificação que explore o facto de que há valores mais frequentes que outros, influencia a codificação a usar. Podem-se aplicar códigos mais curtos a estes do que aos que surgem esporadicamente tal como sucede na codificação de Huffman.
- ? **Exploração do sistema visual humano**[1 e 2]: A visão humana, consumidora final de todo o processo de codificação de vídeo, tem algumas particularidades que podem ser aproveitadas para reduzir a informação necessária como por exemplo:
 - o Limitada resolução espacial – O ângulo de visão é limitado e, consequentemente, o tamanho das imagens a apresentar ou o nível de detalhe das mesmas pode ser adaptado pois há limites ou detalhes que o olho não detecta;
 - o Número de imagens por segundo da visão – Como o olho do ser humano não detecta variações rápidas de imagens acima de um determinado limiar, existe um limite para a frequência temporal das imagens a apresentar sem perda de qualidade perceptível;
 - o Noção de qualidade de imagem – embora seja um critério subjectivo e variável para cada um de nós, existe um limite que satisfaz o receptor de uma imagem de vídeo. Na prática, esta questão leva à definição de valores mínimos da relação sinal ruído de uma transmissão de vídeo.
 - o Como a nossa visão é mais sensível à luminância do que à crominância, várias normas seguem uma estrutura em que a componente da primeira é 4 vezes superior à da segunda pois as amostras de luminância são duas vezes mais na horizontal e na vertical da imagem.

Desde os primórdios da codificação de vídeo que todos os sistemas tentam explorar as condições acima descritas para minimizar a quantidade de informação necessária a uma transmissão de vídeo eficaz.

Nos pontos seguintes analisamos as técnicas fundamentais para a eliminação de redundância em codificação de vídeo e que são a base de quase todos os sistemas actuais.

II.1.1 Compensação de Movimento

A Compensação de Movimento é uma técnica quase universal em sistemas de compressão de vídeo. Entre duas imagens consecutivas, muitos dos pixels mantêm-se inalterados, ou seja, existe redundância temporal entre as imagens consecutivas de um trecho de vídeo. Surgiu, assim, a ideia de reduzir a quantidade de informação a enviar para as imagens seguintes através da supressão desta informação redundante. Se, para a construção de uma nova imagem, partindo do conhecimento prévio da anterior, apenas tratarmos a informação dos pixels alterados, pode-se facilmente construir a imagem seguinte a partir de uma quantidade mais pequena de informação.

Por utilizar informações entre imagens (*frames*) diferentes, este método pode ser classificado por Predição Interframe.

Através de recurso a um exemplo concreto, vamos descrever o processo. O sistema codificador armazena a informação referente a uma imagem. Esta informação é comparada com a imagem seguinte (Fig. II-1a):



Fig II-1a – Duas imagens consecutivas de um trecho de vídeo [29]

A informação respeitante a todos os pixels que se mantiverem não será enviada para a construção da nova imagem. Este primeiro passo já significa uma grande redução de quantidade de informação que teria de passar pelo canal de largura de banda limitada (Fig II-1 b):



Fig II-1b – Diferença entre as duas imagens consecutivas exibidas na fig. II-1a [29]

A partir, apenas, do residual entre as duas imagens, para trabalhar as zonas em que os pixels não se equivalem, a imagem é dividida em macroblocos, tipicamente de 16x16 [1 e 3]. A informação de cada macrobloco é comparada com áreas vizinhas na imagem anterior. Tenta-se encontrar a zona em que as diferenças com o macrobloco em estudo sejam menores para ser determinado um vector de movimento, ou seja, a direcção e distância percorrida pela nossa porção de imagem (Fig II-1c):



Fig II-1c – Vectores de movimento entre imagens consecutivas [29]

Após este passo, restam apenas as áreas da imagem onde o processo não conseguiu determinar vectores de movimento. No caso apresentado, esta quantidade residual é a seguinte (Fig. II-1 d):



Fig II-1d – Informação residual após o processo de Compensação de movimento [29]

Para os processos seguintes, apenas é enviado o conjunto dos vectores de movimento e esta informação residual. Trata-se de uma enorme redução na quantidade de informação e um passo fundamental de todo o sistema. A redução será tanto maior quanto menor for o movimento da imagem.

A utilização da técnica de Compensação de Movimento implica esforço computacional para que toda a zona vizinha possa ser pesquisada. A área envolvida nesta procura é variável de sistema para sistema, tendo de existir um compromisso entre a capacidade de processamento dos codificadores e a qualidade que se ambiciona.

A seguinte imagem (Fig II-2) mostra um exemplo para uma área da imagem onde se faz a procura de movimento de um determinado macrobloco [27]:

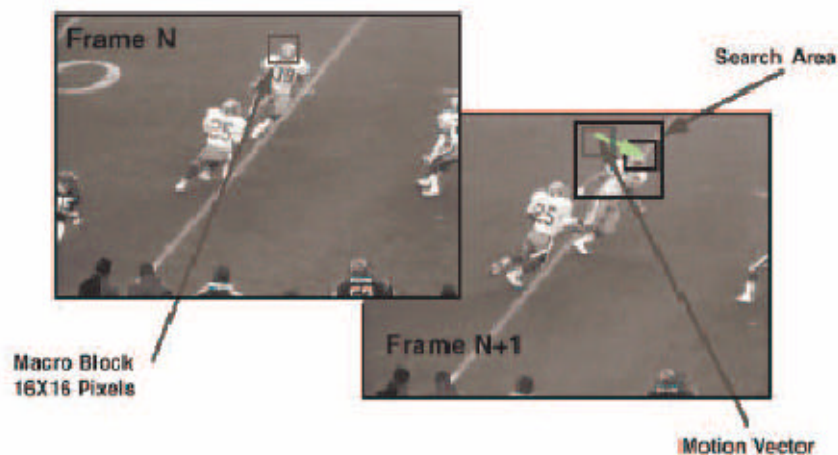


Fig II-2 – Processo de criação do vector de movimento [27]

A implementação prática desta ideia faz-se, ao nível do codificador, através da existência de um decodificador que constrói uma imagem futura adicionando à imagem prévia, guardada num buffer por exemplo, a informação relativa ao tratamento dos vectores de movimento e a compara à imagem que se deveria obter. Obtém-se desta forma uma nova informação a enviar ao decodificador real que é o erro de predição [1].

Um esquema de codificador com Compensação de Movimento encontra-se na seguinte figura (Fig. II-3):

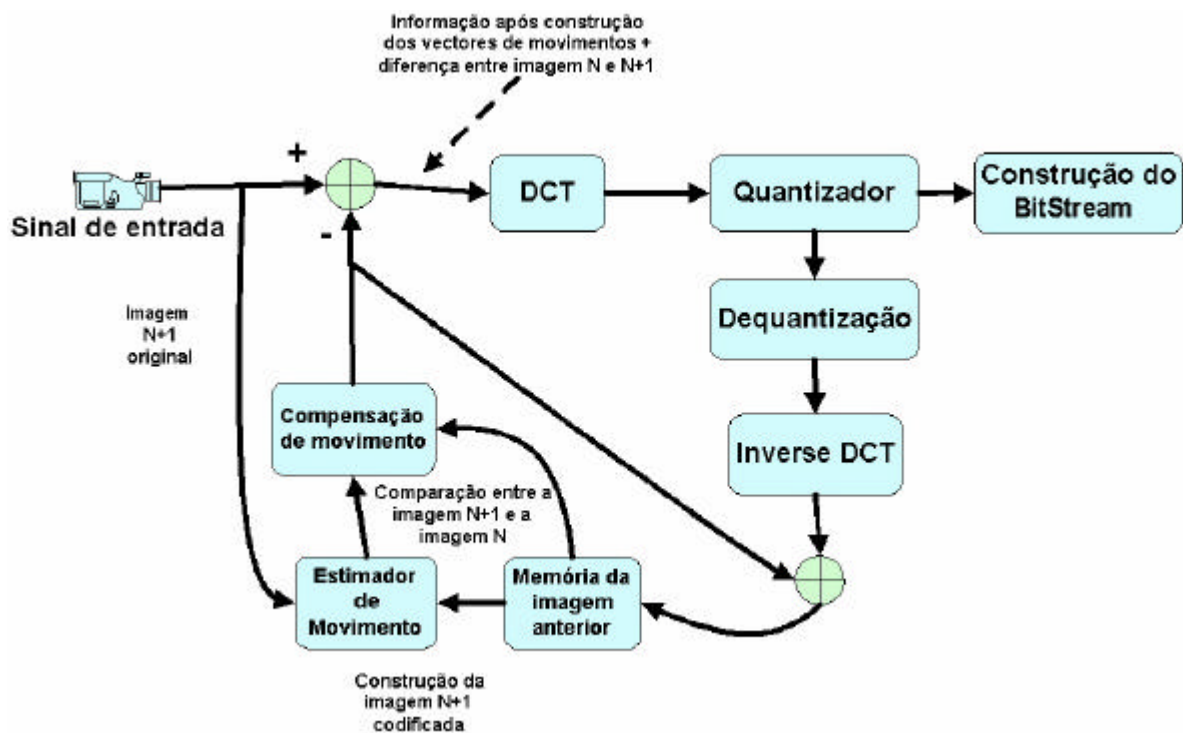


Fig II-3 - Codificador com Compensação de Movimento

II.1.2 Transformada Discreta de Co-seno (DCT)

Outra técnica básica para codificação de vídeo é a Transformada Discreta de Co-seno (DCT). Trata-se de um tratamento matemático da informação relativa a uma determinada imagem do vídeo a codificar, cujo objectivo é a redução da informação por remoção da redundância espacial da mesma. Esta transformada pretende obter a informação relativa a uma porção da imagem no domínio da frequência para permitir um tratamento mais fácil da informação.

A cada bloco de 8x8 pixels é aplicada a transformada, cuja descrição matemática mais comum é a seguinte [6]:

$$C_{m,n} = \frac{1}{4} \sum_{i=0}^7 \sum_{j=0}^7 B_{i,j} \cos\left(\frac{(2i+1)m}{16}\right) \cos\left(\frac{(2j+1)n}{16}\right)$$

Em que:

- ? $C_{m,n}$ representa os coeficientes da transformada
- ? $B_{i,j}$ representa o pixel na linha i e coluna j do bloco de 8x8
- ? $m, n = 0, 1, \dots, 7$
- ? $\frac{1}{4}$ é o factor de normalização
- ? \cos é a função cosseno

O resultado é outro bloco de 8x8 de coeficientes. Cada um destes corresponde a uma contribuição espectral para a imagem. Por exemplo, o primeiro coeficiente diz respeito à componente DC da imagem, ou seja, a sua luminosidade média. Esta transformada não reduz por si só a quantidade de informação. Para isso, o passo seguinte é baseado na

análise estatística de imagens de vídeo, que nos diz que a aplicação da DCT a estas concentra a energia nos coeficientes de baixa frequência [1], o que é um reflexo da alta redundância espacial que se verifica. Muitos dos 64 (8x8) coeficientes têm um valor quase nulo, podendo ser descartados. Como apenas enviamos um número muito inferior aos 64 originais, obtemos uma redução significativa de dados a transmitir.

Uma vantagem importante da DCT é a possibilidade de facilmente se reverter o processo através da sua inversa, a IDCT (Inverse Discrete Cosine Transform) cuja fórmula é a seguinte:

$$B_{i,j} = \frac{1}{8} \sum_{m=0}^7 \sum_{n=0}^7 C_{m,n} \cos\left(\frac{(2m+1)i}{16}\pi\right) \cos\left(\frac{(2n+1)j}{16}\pi\right)$$

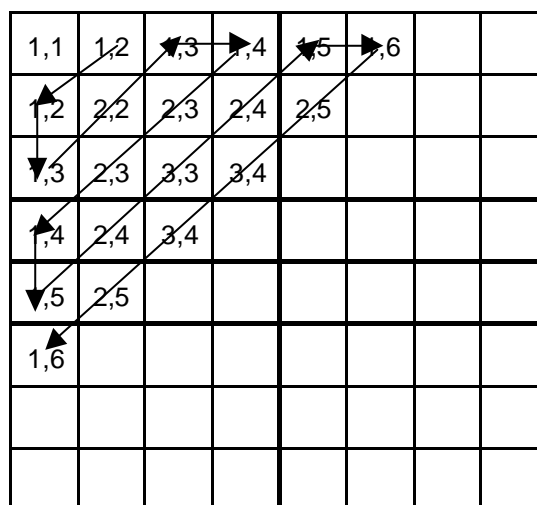
Em que:

- $\frac{1}{8}$ $\frac{1}{8}$ $\frac{1}{8}$ $\frac{1}{8}$ $\frac{1}{8}$ $\frac{1}{8}$ $\frac{1}{8}$ $\frac{1}{8}$
- $\frac{1}{4}$ $\frac{1}{4}$ $\frac{1}{4}$ $\frac{1}{4}$ $\frac{1}{4}$ $\frac{1}{4}$ $\frac{1}{4}$ $\frac{1}{4}$
- i, j 0 e i, j 7
- $C_{m,n}$ representam os coeficientes da transformada
- $B_{i,j}$ representa o pixel na linha i e coluna j do bloco de 8x8

II.1.3 Quantificação e Codificação por VLC

Para melhorar o resultado obtido após a DCT, costuma combinar-se esta com técnicas apropriadas para a Quantificação: os coeficientes escolhidos são quantificados, ou seja, transformados em informação digital que apenas pode ter um determinado número de valores, ou seja, níveis. A escolha do número de níveis de quantificação será tanto maior quanto maior for a qualidade da imagem reproduzida. No entanto, deve-se dimensionar um sistema deste tipo como um compromisso entre a qualidade e a disponibilidade do canal de transmissão.

A construção do bitstream final é feita através do varrimento em Zig-zag desde o coeficiente seguinte à componente DC (coeficiente 1,1 na Fig. II-4) até ao inferior direito:



O coeficiente no topo superior esquerdo corresponde à componente DC, ou seja, a energia média ou luminância da imagem e é tratado à parte.

Como, após a quantificação, muitos dos coeficientes são nulos, dão-se longas cadeias de zeros. Em vez de se transmitir todos estes zeros, envia-se apenas a sua contagem, $r0$, antes do seguinte coeficiente não-nulo, c , através de pares $\{r0, c\}$, por forma a reduzir a entropia do sistema. A codificação destes pares é feita com Códigos de Comprimento Variável (VLC).

Finalmente, como este tipo de codificação gera sequências aleatórias com taxas de bits variáveis, costuma incorporar-se um *buffer* para que a informação a enviar ao canal de transmissão seja mais regular.

Um esquema muito simplificado de um codificador clássico com as técnicas atrás descritas é o exibido na figura seguinte (Fig. II-5):

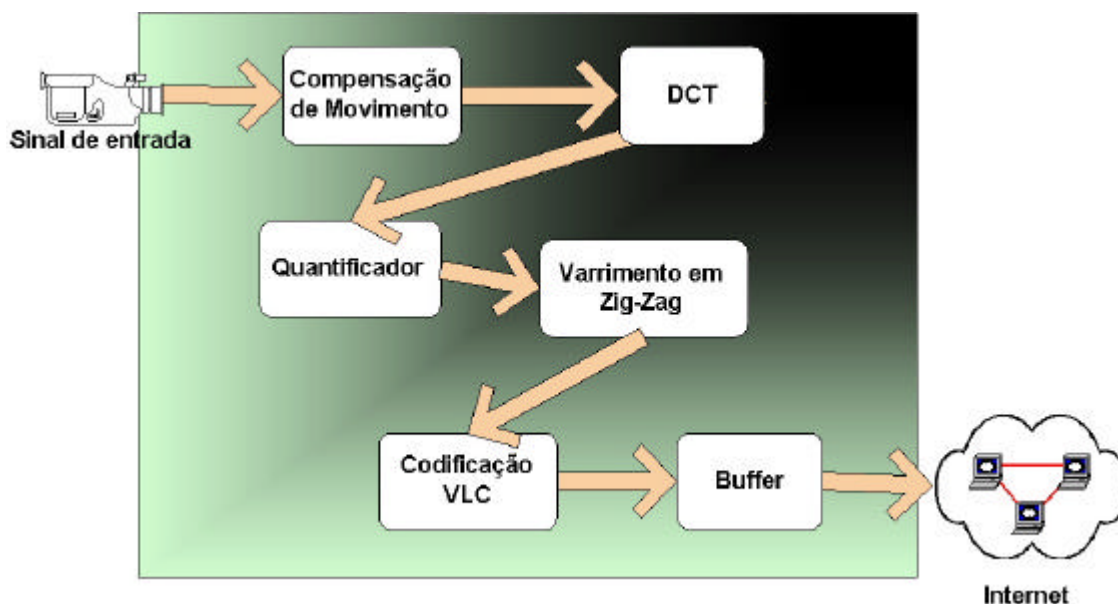


Fig II-5 – Codificador Clássico com Compensação de Movimento, DCT e Quantificação

II.1.4 - Organização do bitstream

À saída do codificador, teremos um fluxo de bits único que inclui o vídeo e o áudio e que é a soma de fluxos de bits elementares. Cada um destes tem uma estrutura hierárquica composta por ordem crescente de complexidade :

- ? **Bloco:** é a quantidade mais pequena de informação à qual pode ser aplicada uma DCT e que corresponde a um conjunto de 8x8 pixels.
- ? **Macro-blocos:** resultantes da informação relativa à compensação de movimento e quantificação de um bloco de 16x16 pixels (4 blocos).
- ? **Fatias:** divisões dos quadros e que resultam da agregação de macro-blocos
- ? **Quadros:** são o equivalente a uma imagem. Há quadros de 4 tipos:
 - o I : quadros codificados apenas com a informação do próprio (intra-quadro). Por exemplo, na norma MPEG-1, em cada sequência de 132 quadros pelo

menos um tem de ser deste tipo pois é aqui que é feito processo para acesso aleatório.

- P: codificados com informação dos quadros anteriores (inter-quadro)
 - B: codificado com informação de quadros anteriores e/ou posteriores
 - D: codificados unicamente com a componente DC de cada bloco
- ? **Grupo de quadros**: Conjunto de quadros e que permitem o acesso aleatório a uma parte do vídeo
- ? **Sequências**: conjuntos de vários grupos de quadros e que são um trecho de vídeo independente.

Resulta da definição de quadro tipo B, que a transmissão dos quadros pode não ser igual à ordem natural dos mesmos. Por isso mesmo, o decodificador terá de interpretar as informações que circulam associados aos fluxos e que lhes dão as indicações de localização temporal dos mesmos.

Deve referir-se que cada norma tem a sua própria forma de implementar o bitstream e, por exemplo, pode nem sequer utilizar todos os tipos de quadros ou nem sequer ter fatias.

II.2 Primeiras Normas de Codificação de vídeo: H.261, H.263 e MPEG-1 e MPEG-2

O desenvolvimento de qualquer tipo de tecnologia, quando há múltiplos investigadores e implementadores interessados, apenas será eficaz se se estabelecerem regras ou normas que indiquem os caminhos a tomar em conjunto. A transmissão de vídeo e a codificação associada não foge à regra e foram criadas várias normas que acompanharam a evolução da tecnologia. Em seguida serão apresentadas as mais importantes.

II.2.1 H.261

Quando ainda se designava por CCITT, em 1990, a actual ITU estabeleceu uma recomendação para adaptar a transmissão de vídeo à então emergente tecnologia RDIS.

A norma H.261 pertence à família H.320 para transmissão de vídeo-telefone e vídeo-conferência e foi definida para operar a débitos múltiplos de 64Kbps ao que corresponde um múltiplo inteiro de canais RDIS, desde 1x64 até 30x64 (1,92 Mbps). Por isso, também se costuma designar o H.261 por *px64*.

A norma combina técnicas *intraframe* e *interframe*, por forma a eliminar as redundâncias espaciais e temporais, respectivamente. Esta norma utiliza, tipicamente, os seguintes parâmetros:

- ? Compensação de movimento em macroblocos 16x16, com pesquisa de movimento numa área de até 15 pixels de distância [2].
- ? Aplicação da DCT a blocos de 8x8 pixels.
- ? Quantificação dos dados combinada com um *buffer* para regulação da taxa de transmissão.
- ? Possibilidade de eliminar quadros de imagem em caso de congestionamento.

Para garantir a funcionalidade na gama mais baixa, a 64 Kbps, é necessária uma grande taxa de compressão, chegando esta, por vezes, a ser de 2000:1 [4]. Com este nível de compressão, imagens com movimentos bruscos ficam seriamente afectadas e por isso, apenas se utiliza para aplicações mais simples como o vídeo-telefone. Por outro lado, na sua máxima capacidade ($30 \times 64 = 1,92$ Mbps) a qualidade é semelhante à de um vídeo VHS [2].

O H.261 utiliza formatos específicos para o tamanho das imagens:

- ? Common Intermediate Format (CIF) – 288x356 (linhas x colunas) [2].
- ? Quarter Common Intermediate Format (QCIF) – melhor adaptado para menores taxas de transmissão, tem um tamanho de imagem de 144x160;

Para cada um destes formatos, as frequências de varrimento possíveis são 30 Hz, 15Hz, 10 Hz ou 7.5 Hz.

O H.261 foi a primeira norma que se impôs como um verdadeiro standard de compressão de vídeo, sendo considerada a base da tecnologia de Compressão de vídeo [5]. Foi progressivamente ultrapassada pela norma H.263 que estudaremos em seguida.

II.2.2 H.263

Após a bem sucedida implementação do H.261, o ITU-T tentou expandir a aplicação destas técnicas a velocidades de transmissão mais baixas, dado o crescimento de formas de acesso com recursos limitados como os primeiros modems analógicos que operavam a velocidades inferiores a 64 Kbps, primeiro patamar do H.261. Para alcançar este objectivo foi especificada pelo ITU, em 1995, uma nova norma, denominada H.263, que se baseou nos trabalhos prévios da norma H.261.

Assim, a estrutura básica de um sistema de codificação de vídeo em H.263 é muito semelhante ao do seu antecessor e tem as seguintes características gerais:

- ? Para a construção de uma imagem não são enviadas informações sobre os pixels que se mantêm da imagem anterior;
- ? Utiliza a compensação de movimento para reduzir a quantidade de informação relativa a partes da imagem que se moveram, utilizando os já descritos vectores de movimento;
- ? A busca de movimento de um macrobloco é limitada a um espaço de distância em relação à posição original inferior a 16 ou 31.5 pixels, consoante o modo de predição escolhido, para limitar o esforço computacional exigido [6];
- ? Permite que os vectores de movimento apontem para fora da área da imagem, o que melhora a predição de movimentos nas fronteiras da mesma.
- ? Elimina as redundâncias espaciais através da transformada DCT melhorada;
- ? A quantificação foi definida para permitir até 62 níveis de quantificação positivos e outros tantos negativos mais o nível 0 que representa a componente DC [6];

Outras inovações introduzidas foram as seguintes:

- ? Estrutura da imagem permite novos formatos [6]:

- Sub-QCIF
 - QCIF
 - CIF
 - 4CIF
 - 16CIF
- ? O movimento entre imagens de cada macrobloco pode ser descrito por 4 vectores de movimento em vez de um único, o que confere maior precisão na localização do macrobloco na nova imagem;
 - ? Possibilidade de fazer a estimação do movimento em meias unidades elementares da imagem (pixels);
 - ? Codificação por VLC após a construção do bitstream quantificado;
 - ? Permite a aplicação de codificação aritmética, o que reduz em 5% o tamanho do bitstream em relação à codificação por códigos de Huffman;
 - ? Utiliza quadros PB, ou seja blocos com predição com base na imagem anterior (P) intercalados com imagens com predição bidireccional (B) o que implica um duplicar da quantidade de imagens transmitidas mantendo a taxa de transmissão.

A eficiência de codificação do H.263 é superior à do H.261 em cerca de 1.5 a 2 vezes [4], o que significa que o bitstream final tem, na melhor hipótese, um tamanho ligeiramente inferior a metade do que na anterior norma.

Um codificador H.263 típico é o seguinte [7] (Fig. II-6):

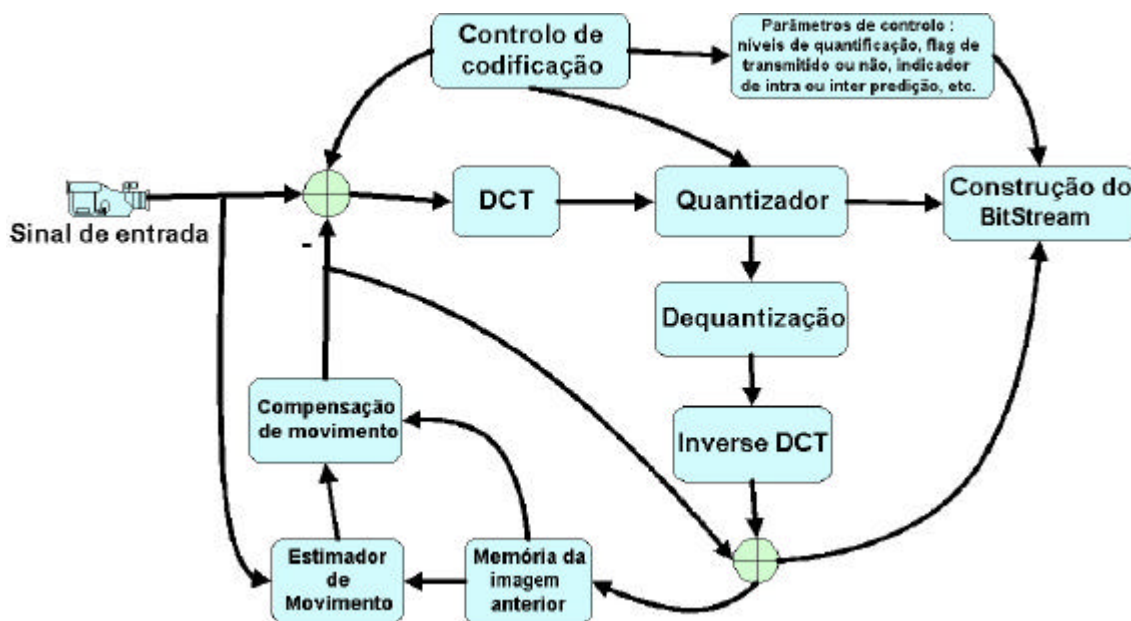


Fig II-6 - Codificador típico para H.263

Esta norma teve 2 desenvolvimentos que melhoraram as suas características e que foram denominadas H.263+ e H.263++.

H.263+

A norma H.263+ é uma evolução da H.263 base com 12 novos modelos opcionais [6] descritos nos anexos D, e do I ao T do documento do ITU-T. Entre eles destacam-se as seguintes novidades [6]:

- ? Implementa os denominados Reverse VLC's. Nestes, a descodificação com códigos de comprimento variável pode ser feita tanto na direcção de leitura como na via inversa. Ou seja, se ocorrer um erro de transmissão a informação dos vectores de movimento, por exemplo, não se perde pois o descodificador pode obtê-los na via inversa. Mesmo nos vectores há evolução: agora podem apontar até uma distância de -256 ou + 255.5 pixels.
- ? Aplica filtros na predição das imagens para proporcionar redução de erros de reconstrução nas fronteiras dos blocos.
- ? Divisão da imagem em fatias que têm um número variável de macroblocos para que partes da imagem correlacionadas possam ser tratadas como um todo.
- ? Predição de blocos a partir de informação recolhida em macroblocos vizinhos.
- ? Introdução de informação adicional no bitstream para mais funcionalidades de apresentação do vídeo;
- ? Possibilidade de predição da imagem seguinte a partir de outras imagens que não só a anterior para impedir a propagação de erros se uma imagem contiver erros;
- ? Possibilidade de suportar escalabilidades temporais, espaciais ou de SNR. Definido no anexo O, este modo especifica a sintaxe para suportar esta característica fundamental para a adaptação a diferentes condições de transmissão para diferentes utilizadores como é o caso da Internet. Dada a sua importância para a transmissão de vídeo na Internet, estas técnicas serão mais aprofundadas no ponto III.2.
- ? Modo de quantificação Modificado: permite a introdução de métodos de controlo do tamanho do bitstream tais como a alteração do número de níveis de quantificação, maior precisão nos níveis de crominância ou aumento do número de coeficientes da DCT.

Cada um dos modos pode ser utilizado isoladamente ou combinado com outros, o que faz com que o H.263+ seja muito adaptável às diferentes condições ou qualidades dos codificadores, descodificadores ou mesmo dos meios de transmissão. Para que existisse alguma standardização, o ITU-T estabeleceu alguns níveis que não são mais que conjuntos de modos. No que diz respeito ao assunto em estudo neste trabalho, esta flexibilidade faz com que o H.263+ seja um dos primeiros métodos de codificação e transmissão de vídeo com alguma capacidade para responder aos requisitos da transmissão na Internet.

H.263++

A versão final da norma H.263 deu-se com a sua evolução H.263++ (ou versão 3) finalizada em Julho de 2000 [5]. As suas condições estão representadas nos Anexos U, V e W e adicionam ainda mais 7 modos aos 12 já introduzidos no H.263+.

As principais características adicionais são [5, 8, 9]:

- ? Possibilidade de ser identificada uma implementação específica para a IDCT.
- ? Introdução de mensagens no bitstream acerca do conteúdo, como por exemplo, indicações de direitos de autor, texto avulso, descrição do vídeo, etc..

- ? Implementação de um sistema, denominado Enhanced Reference Picture Selection (ERPS) que permite múltiplas referências a imagens anteriores na predição da mesma imagem. Por exemplo, na predição de uma imagem, um vector de movimento pode ser construído com referência a um macrobloco de uma imagem que não é a referência utilizado por outro vector da mesma imagem.
- ? Associado ao ERPS existe a possibilidade de apenas serem indicadas áreas parciais das imagens anteriores para referência na construção da seguinte de forma a reduzir a quantidade de memória necessária nos codificadores e decodificadores.
- ? Existência de mensagens de realimentação enviadas do decodificador para o codificador para identificar imagens ou partes de imagens mal codificadas.
- ? Maior resistência a erros através da partição de dados no envio do bitstream.
- ? Reforça a resistência a erros de propagação através do aperfeiçoamento da técnica dos Reverse VLC's já explicada no H.263+.

Em comparação ao codificador H.263 base, um codificador capaz de suportar H.263++ apresenta a seguinte configuração (Fig II-7):

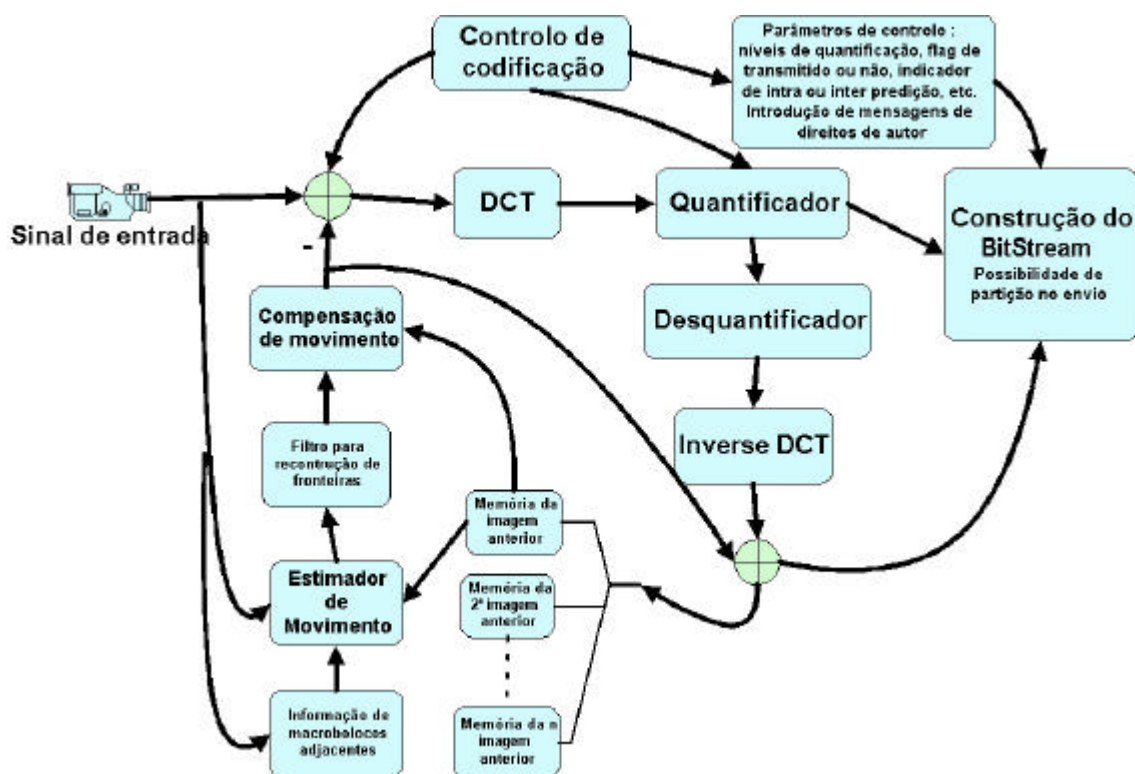


Fig II-7 – Codificador típico para H.263++

Pode-se verificar na figura a existência de memória para várias imagens anteriores.

Podemos considerar o H.263 como a primeira norma de codificação de vídeo com objectivos de transmissão na Internet. Primeiro, porque foi desenvolvida para suportar baixas taxas de transmissão, como aquelas que maioritariamente constituíam e continuam ainda a constituir a maioria dos acessos dos consumidores de vídeo na Internet. Depois,

porque introduziu uma série de desenvolvimentos para melhor se comportar em situações frequentes na rede, como por exemplo a perda de pacotes ou os erros de transmissão. E, também, porque incorpora conceitos básicos de escalabilidade para poder adaptar o seu conteúdo às diferentes condições de cada utilizador.

II.2.3 MPEG-1 e MPEG-2

O MPEG, desenvolvido pelo Motion Picture Experts Group, é um conjunto de normas que define a codificação de imagens de áudio e vídeo e a sua compressão num formato digital. Esta norma permitiu a criação de ficheiros que tinham uma melhor qualidade do que outros formatos de codificação para o mesmo grau de compressão, e tornou-se de longe o método mais usado para estes fins. Ao contrário de outros formatos como o Windows Media ou o Real Player que são proprietários, o MPEG é um código aberto, permitindo contribuições de múltiplas entidades no seu desenvolvimento.

A evolução da norma ao longo dos últimos anos tem vindo a aperfeiçoar as técnicas de compressão e adaptação aos meios de transporte e tem sido dada grande importância à sua implementação em meios com baixas taxas de transmissão ou que proporcionem pouca qualidade de serviço como o caso da Internet.

O início das investigações remonta a 1988 e a primeira norma definitiva, o MPEG-1, em 1992, criou a forma de codificação que viria a ser usada nos CD-ROMs de vídeo e no MP3[2]. Permite a codificação a uma taxa de 1,5 Mbps, o que dificulta a sua utilização na Internet pelas razões apontadas no Capítulo I.

A resolução máxima da imagem após codificação é de 576 x 768 a uma frequência de varrimento de 30 Hz.

Basicamente, a codificação e compressão MPEG utiliza as técnicas já anteriormente descritas desde a compensação de movimento até aos códigos de comprimento variável. Prevê a utilização de três tipos de quadros: I, P e B. Os quadros do tipo D apenas existem quando se pretendem funções de procura rápida [2] .

A norma MPEG-1 define como MPEG Stream uma sequência de bits transmitida e que pode conter até 32 fluxos de áudio e 16 de vídeo mais 2 fluxos de uso para referências e outras aplicações.

O MPEG-1 estabelece ainda a forma como é feita a sincronização destes fluxos indicando o tipo de quadros a usar, permite o acesso aleatório, determina o tipo de compressão e a gestão de buffers.

O MPEG-2 representou uma importante evolução em relação ao MPEG-1, muito por causa do crescimento das tecnologias de vídeo e das próprias redes de transmissão que passaram a possibilitar muitos tipos de acesso e a velocidades mais rápidas.

Muitas das definições do MPEG-1 continuaram válidas no MPEG-2. As alterações introduzidas pretenderam suportar novas funcionalidades ou tecnologias [2]:

- ? Transmissão digital da televisão nos formatos PAL, SECAM ou NTSC
- ? DVD
- ? Aplicações multimédia como a televisão digital de alta definição
- ? Taxas de transmissão até 100 Mbps
- ? Diferentes resoluções espacio-temporais

Podemos fazer as seguintes comparações entre os 2 formatos:

- ? A norma MPEG-2 não suporta os quadros tipo D, mas mantém as definições dos outros três tipos.
- ? A norma MPEG-2 passou também a permitir a codificação de imagens no modo entrelaçado.
- ? Evoluindo em relação ao MPEG-1, a norma MPEG-2 suporta 4 níveis:
 - o baixa resolução (352x240)
 - o resolução média ou Principal (720x480)
 - o resolução alta-144 (1440x1152)
 - o resolução alta (1920x1080).
- ? Além do áudio e vídeo que formavam o sinal MPEG-1, a norma MPEG-2 possibilita um número quase ilimitado de trechos de dados.

Para conseguir esta evolução utiliza-se uma diferente forma de construir as bitstreams, nomeadamente:

- ? Os macro-blocos continuam a ser conjuntos de 4 blocos no que diz respeito à luminância, mas permite 3 tipos de sub-amostragem da crominância o que resulta em 3 tipos diferentes de macro-blocos (6, 8 ou 12 blocos);
- ? Codificação intra-quadro dos macro-blocos com códigos diferentes e com a possibilidade de divisão da componente DC por 8,4,2 e 1 e a componente AC codificada com coeficientes de 2048 degraus de quantificação;
- ? Possibilidade de utilização do modo entrelaçado, em que a sequência de bits transporta alternadamente subquadros ou linhas pares e ímpares, o que diminui a sensação cintilante na visão humana;
- ? Novos padrões no varrimento em ziguezague dos coeficientes DCT;

Tal como em todas as normas MPEG, apenas é definido o formato do bitstream final e não a forma de codificação [28]. A sequência binária obtida agrega vários algoritmos de compressão adaptados a diferentes situações mas permite que dela sejam retiradas as diferentes versões do mesmo vídeo, cada uma com diferentes níveis de qualidade.

É a primeira aplicação de técnicas de compressão escaláveis em que a uma camada base, capaz de apresentar uma versão do vídeo com o mínimo de qualidade, se pode adicionar outras camadas que vão melhorando a qualidade da apresentação, tal como será descrito no Capítulo III nas técnicas de compressão adaptadas a *streaming* de vídeo.

A norma MPEG-2 criou uma forma de permitir graus de liberdade aos utilizadores em relação àquilo que se pretende obter de uma determinada sessão de *streaming*, com a criação dos denominados Perfis. Cada perfil define as condições necessárias ao nível dos codificadores, descodificadores e algoritmos de compressão para uma sessão em MPEG-2 de acordo com os objectivos. Assim, se se pretender, por exemplo fazer apenas uma simples sessão de videoconferência com qualidade reduzida, não é necessário implementar toda a complexidade necessária para um *streaming* de áudio e vídeo para difusão de televisão digital e assim, abrimos a possibilidade de acesso a sessões com qualidade limitada de utilizadores com poucos recursos. Os perfis definidos pelo MPEG-2 são [2]:

- ? **Perfil Simples:** indicado a aplicações que exijam pequenos tempos de atraso. Não são utilizados os quadros do tipo B e assim não exige reordenação das imagens e utiliza bitstreams mais leves. Apenas permite o nível principal;
- ? **Perfil Principal:** Utiliza os quadros do tipo B, aumentando a qualidade mas exigindo boa capacidade de transmissão para evitar os atrasos. Utiliza os 4 níveis e permite débitos desde 4 Mbps até 80 Mbps;
- ? **Perfil Escalonável em SNR :** Suporta ajustes para diferentes SNR e quantificação de coeficientes DCT para permitir fornecimento de serviço em diferentes situações. Estamos perante um perfil escalável, com camada base que pode começar em 3 Mbps.
- ? **Perfil Escalonável Espacial:** Suporta ajustes para diferentes resoluções espaciais. Utiliza somente o nível alto 1440 e uma taxa de 60 Mbps. A camada base por si só requer 15 Mbps;
- ? **Perfil Elevado:** Representa o topo de qualidade do MPEG-2, exigindo a amostragem de vídeo mais qualitativa. Se for utilizado o nível alto produz uma bitstream a 25 Mbs.

Os codecs que funcionam para os perfis mais exigentes têm de possuir a capacidade de interpretar as bitstreams dos perfis mais baixos. Um esquema simplificado de um codificador para MPEG-2 poderia ser o seguinte (Fig II-8):

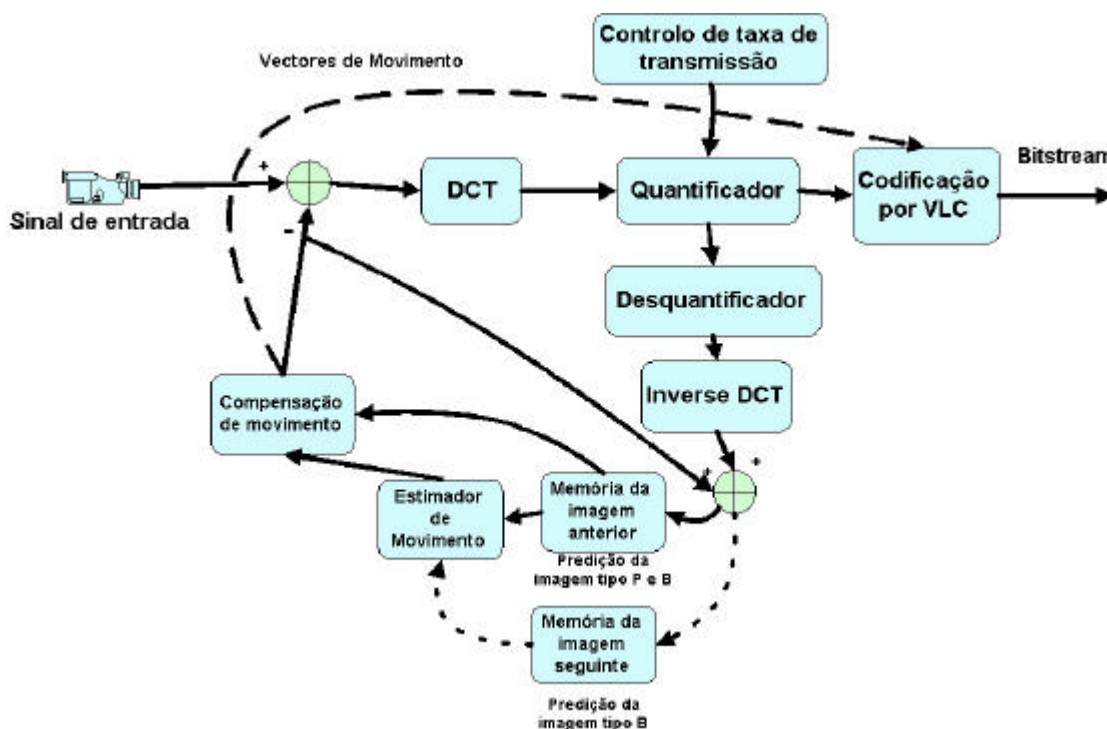


Fig II-8 - Codificador típico para MPEG-2

Embora o objectivo do MPEG-2 fosse aumentar os débitos binários do vídeo codificado, as características atrás referidas permitiram também a implementação desta técnica ligada a utilização na Internet, principalmente porque introduz características de escalabilidade, embora não muito sofisticadas, relacionadas com a qualidade de serviço. Contudo, estamos

a lidar com débitos binários muito altos para a esmagadora maioria das ligações utilizadas mesmo no caso de empresas. Para tentar expandir a técnica de compressão a utilizadores mais limitados, o MPEG Group voltou a reunir esforços para desenvolver uma norma que possibilitasse o *streaming* de vídeo em débitos mais baixos, com o objectivo de o adaptar à Internet.

II.3 - MPEG-4

O MPEG-4 é um standard desde 1999 que surgiu na sequência dos desenvolvimentos da norma MPEG com o objectivo de ultrapassar as limitações dos formatos anteriores, nomeadamente o MPEG-1 e MPEG-2, que estavam orientados à codificação por pixel/imagem e que tinham atingido o limite de desenvolvimento. Para responder à progressão das redes móveis e à necessidade de expandir as técnicas de compressão de vídeo a meios com baixas taxas de transmissão de dados foram desenvolvidas várias investigações que resultaram no MPEG-4.

O MPEG-4 foi o responsável por um importante avanço na área da compressão de vídeo e contribuiu para a convergência entre tecnologias tais como a televisão e a informática.

Os principais objectivos desta norma eram [26]:

- ? Ser facilmente adaptada aos protocolos de rede dos operadores.
- ? Permitir o acesso à tecnologia de utilizadores ligados a baixas taxas de transmissão.
- ? Permitir uma maior interactividade e serviços mais completos aos receptores da informação.
- ? Proteger de forma mais eficaz os direitos de autor dos produtores.

Para além de alcançar estes objectivos, o resultado deste desenvolvimento foi a abertura de novas possibilidades para pesquisa e acesso a bases de dados, mais fácil edição e manipulação dos conteúdos, melhor acesso aleatório, maior eficiência na compressão e maior escalabilidade no tipo de ligação efectuada o que lhe permite ser transparente ao tipo de rede que a suporta.

As diferenças principais em relação aos seus antecessores são:

- ? No MPEG-1 e 2, as codificações de áudio e vídeo eram consideradas ao nível do transporte como uma única entidade, ou seja, tinham ambas o mesmo endereçamento IP e a separação só podia ser feita pelo decodificador. No MPEG-4, são enviadas em diferentes portas IP, o que permite por exemplo que um receptor possa optar por apenas receber a componente visual de um ficheiro, melhorando a rapidez na transmissão [22].
- ? Com o MPEG 1 e 2, a representação da imagem é feita com base na informação da imagem considerada como um todo. No MPEG-4 podem ser diferenciados os vários elementos da cena (pessoas, objectos, etc.) quer no espaço quer no tempo. Ora isto permite tratar cada um dos componentes desta imagem e, se assim for autorizado, possibilitar ao receptor obter diferentes versões da imagem transmitida quer por alteração dos parâmetros de visualização (mudar o ângulo de visão por exemplo) quer através da manipulação destes objectos (mudar o local de objectos, retirar outros, etc.).

O Forum MPEG definiu para o desenvolvimento do MPEG-4 uma estrutura em partes, cada uma com áreas concretas de actuação embora sujeitas a regras comuns:

1	Sistemas
2	Visual
3	Áudio
4	Teste de Conferência
5	Software de Referência
6	Adaptação à camada de transporte - DMIF (<i>Delivery Multimedia Integration Framework</i>)
7	Software Optimizado para ferramentas MPEG-4 (codificadores)
8	MPEG4 em IP
9	Descrição de Hardware Referência
10	Norma AVC (<i>Advanced Video Coding</i>)

II.3.1 - Modo de Funcionamento

O MPEG-4 estabeleceu uma ruptura com os seus antecessores na forma de tratamento de uma imagem pois a codificação de um vídeo resulta da representação individual dos elementos presentes. Esses elementos podem ser objectos físicos, como por exemplo mobílias, paredes e pessoas ou elementos de áudio que façam parte da cena captada. Podem ainda surgir elementos introduzidos artificialmente como caracteres inseridos para comentários, legendas, gráficos, etc.

Cada um destes elementos, denominado “*Media Object*” (no MPEG-4 Visual é denominado *Video Object* – VO) é tratado individualmente sendo-lhe referenciadas condições temporais e espaciais que permitirão ao decodificador reconstruir a imagem [26]. A norma permite a associação de *media objects* de forma hierárquica para criar objectos compostos que são posteriormente multiplexados e sincronizados para transmissão sobre um canal que forneça os requisitos de QoS apropriados. Um exemplo de objecto composto é o conjunto formado por dois objectos primitivos: um ser humano e a sua voz (Fig II-9).

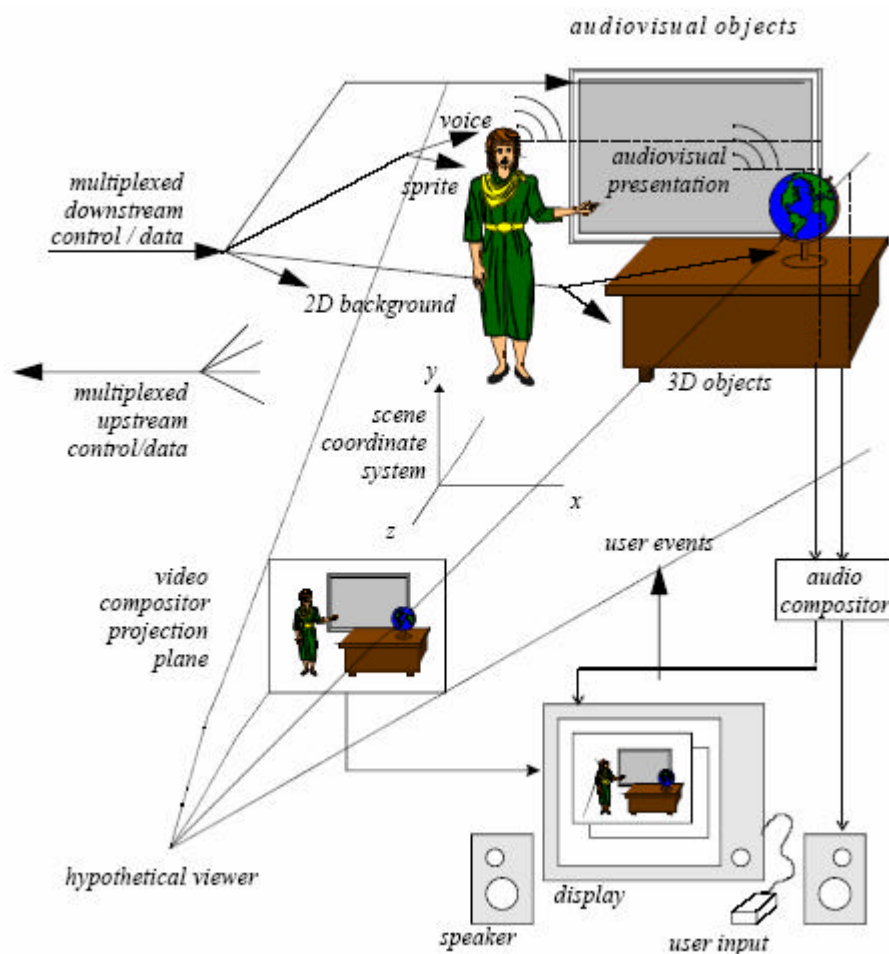


Fig II-9 - Decomposição da imagem em Video Objects (VO's) [26]

A representação codificada de um *media object* é feita por um conjunto de dados elementares que possuem informações úteis para o decodificador reconstituir a imagem, como por exemplo, as coordenadas da sua posição original na cena. A coordenação destas sequências de bits é feita através do *object descriptor*, que indica os trechos que constituem o objecto representado e também as necessidades de Qualidade de Serviço para a sua transmissão. A sincronização dos trechos é feita pela chamada camada de sincronização, guiando-se através de marcas temporais associados a entidades de acesso de cada uma, como por exemplo os quadros de vídeo.

Em conjunto com a camada de sincronização, o MPEG-4 define a camada de transporte para estabelecer as regras de transporte dos trechos. Esta camada é constituída por duas subcamadas multiplexadas, uma das quais define o agrupamento dos trechos elementares e para a qual foi dedicada a parte 6 do MPEG-4, denominada *Delivery Multimedia Integration Framework* (DMIF), na qual foi definida a ferramenta *FlexMux* cujos objectivos são o de simplificar e emagrecer os recursos alocados à entrega através da redução do número de ligações na rede e do atraso ponta a ponta. Outra subcamada é responsável por fornecer o transporte adequado à Qualidade de Serviço definida para a sessão de vídeo. Somente a interface com esta subcamada é especificada implicitamente pelo MPEG-4, o que foi feito intencionalmente para permitir que qualquer protocolo de

transporte se possa adaptar à transmissão de ficheiros de MPEG-4. Por exemplo a pilha protocolar RTP+UDP+IP, ou o AAL5 do ATM adaptaram-se a esta funcionalidade [26]. O receptor pode agir sobre os elementos recebidos de acordo com graus de liberdade pré-definidos, tais como colocar objectos num qualquer ponto de um sistema de coordenadas, aplicar mudanças à forma, reagrupar objectos compostos noutros mais complexos ou alterar o ângulo de visão de uma cena. Este aspecto levanta questões de propriedade intelectual de uma obra de vídeo e o MPEG-4 criou um sistema de identificação de autores através de numeração ou códigos e que acompanham as *streams* de dados.

Os formatos permitidos para vídeo em MPEG-4 variam de 5 Kbps até mais de 1 Gbps, vídeo progressivo ou entrelaçado e resoluções desde sub-QCIF até resolução de alta definição como a utilizada em estúdios profissionais [2 e 26].

Tal como os seus antecessores, existe a diferenciação em perfis subdivididos em níveis consoante o tipo de vídeo e padrões de qualidade exigidos. Cada um dos perfis define a exigência a que cada codificador e decodificador têm de responder e os níveis ajudam a definir a complexidade computacional. Para o MPEG Visual, existem 19 perfis diferentes. Os que mais dizem respeito à transmissão na Internet são:

- ? *Simple Visual Profile* dedicado às redes móveis;
- ? *Simple Scalable Visual Profile* que introduz características de escalabilidade para adaptação a redes heterogéneas;
- ? *Core Visual Profile* que adiciona codificação de objectos ao Simple Visual Profile;
- ? *Core Scalable Profile* que adiciona ao Core Visual Profile a codificação escalável temporal e espacial;
- ? *Advanced Simple Profile (ASP)* que é o perfil que permite, entre outros, compensação de movimento a 1/4 de pixel e frames do tipo B. Surge muitas vezes como a referência para comparações entre o MPEG-4 e outras normas;
- ? *Fine Granularity Scalability Profile* que permite o uso de técnicas desenvolvidas especificamente para melhor escalabilidade, como o FGS (*Fine Granularity Scalability*) que apresentaremos no capítulo III.

II.3.2 Principais funcionalidades para adaptação à Internet

No que diz respeito a este trabalho, merecem especial importância todas as componentes desta norma que facilitam a transmissão de vídeo em formato MPEG-4 pela Internet.

Transporte sobre IP

Como já vimos, não é definida a camada de transporte mas apenas o seu interface. Assim, o transporte sobre IP é possível e existe adaptação específica definida em conjunto com o *Internet Engineering Task Force* (IETF). Foram estabelecidas algumas regras para o transporte baseado em RTP, RSTP ou HTTP entre outros, nomeadamente no que diz respeito à construção dos *payloads* do RTP. Neste caso, as principais características são:

- ? Transporte homogéneo dos vários trechos de MPEG-4;
- ? Correspondência lógica entre os pacotes RTP e os de MPEG-4;
- ? Agregação de vários pacotes MPEG-4 debaixo de um único pacote RTP para minimizar overheads;

- ? Adaptação para transporte de pacotes de trechos elementares associadas pela FlexMux;

DMIF

A parte 6 do MPEG-4 é dedicada a uma componente importante da norma. Cabe à especificação DMIF (Delivery Multimedia Integration Framework) a adaptação entre a aplicação e o transporte de forma a blindar a aplicação à rede de transmissão. O produtor do vídeo deixa de se preocupar com o tipo de rede onde será transportada a sequência de bits pois será a DMIF a adaptar o ritmo de transmissão e a qualidade às características do canal [26]. Está prevista a adaptação a redes tais como IP, RDIS, ATM, POTS ou redes móveis.

Genericamente podemos defini-la como um protocolo de sessão que faz a gestão de *streaming* de multimédia sobre as várias tecnologias de entrega.

O processo pode resumir-se da seguinte forma: aquando do início de uma transmissão, a aplicação do lado emissor pede à DMIF para estabelecer uma sessão com a DMIF correspondente do lado receptor por forma a se criar um caminho entre os 2 pontos. A DMIF receptora indica o pedido à respectiva aplicação que se encarrega de definir canais de comunicação com a aplicação inicial. A DMIF define automaticamente os parâmetros de transmissão e os pontos de comunicação definindo apontadores para os locais onde está a informação que será transmitida através de mensagens de sinalização específicas.

As tarefas mais importantes da DMIF são:

- ? Suporte de todo o tipo de redes: IP, ATM, RDIS, móveis, etc.;
- ? Controlo de Qualidade de Serviço, que pode ser de 3 tipos: controlo contínuo, controlo a pedido ou controlo de violação de QoS;
- ? Suporte de *acknowledgments* entre emissor e receptor;

Compressão de Vídeo para baixos débitos

Tal como noutros formatos, a compressão da informação resultante da captura de vídeo tem por objectivo reduzir ao máximo o tamanho da informação a enviar pela rede para se obter um padrão mínimo de qualidade na imagem reconstituída. A grande diferença no MPEG-4 é que este formato pretende possibilitar o funcionamento de *streaming* de vídeo em ligações de baixo débito o que faz com que este aspecto seja aqui mais importante.

A norma definiu métodos para se obter:

- ? Compressão eficiente de imagens e vídeo
- ? Compressão de texturas para mapeamento a 2 ou 3 dimensões
- ? Acesso aleatório a pontos ou objectos específicos do vídeo
- ? Permissão de manuseamento das imagens
- ? Boa resistência e recuperação a erros
- ? Escalabilidade temporal, espacial e SNR
- ? Escalabilidade com base nas texturas ou no conteúdo das imagens

A codificação de vídeo convencional é feita em moldes semelhantes às utilizadas no MPEG-1 e 2, nomeadamente:

- ? Compensação de movimento
- ? Codificação de texturas
- ? DCT e codificação por VLC

A informação de localização e forma de um objecto é feita pela introdução de informação adicional para codificar estas componentes. Essa informação adicional em formato binário pode utilizar códigos desde um simples bit (por exemplo, 1 para a presença de um objecto e 0 para a sua ausência) até 8 bits em que os 256 estados possíveis podem servir para indicar vários níveis de transparência do objecto, por exemplo.

Cada *Video Object* pode ser codificado utilizando técnicas de compressão escalável que serão explicadas e aprofundadas no Capítulo III:

- ? Escalabilidade espacial
- ? Escalabilidade temporal
- ? Escalabilidade de SNR
- ? FGS (utiliza um máximo de 11 níveis)
- ? Combinação de FGS com escalabilidade temporal

Contudo, a necessidade de servir os utilizadores com baixas taxas de transmissão levou à diferenciação das ferramentas de codificação de vídeo [26] (Fig II-10):

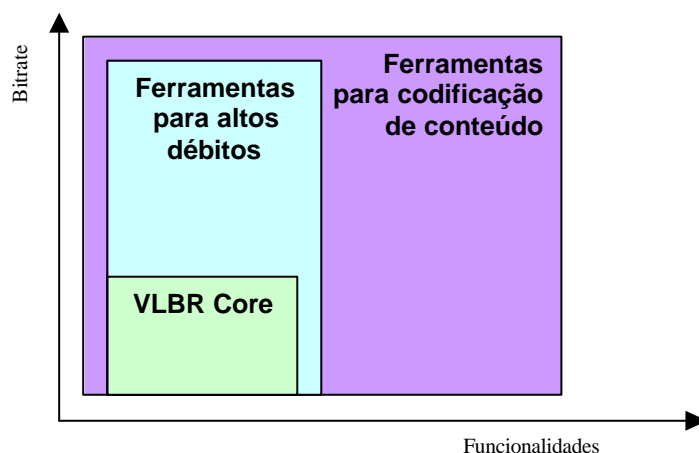


Fig II-10 - Escalões para as ferramentas de codificação no MPEG-4

O *Very Low Bit Rate Core* (VLBR Core) inclui os algoritmos e ferramentas necessários para se produzir *streaming* de vídeo para débitos entre os 5 e os 64 Kbps com varrimento a 15 Hz e baixa resolução espacial, o que cobre as necessidades da maior parte das ligações à Internet do segmento residencial [2]. Esta técnica garante mesmo assim uma eficiente codificação de imagens rectangulares com alta robustez a erros, baixa complexidade e latência, Permitindo aos utilizadores as funções típicas de um gravador de vídeo como o *fast-forward* e acesso aleatório.

Para atingir os escalões superiores, o sistema vai adicionando funcionalidades ao VLBR Core permitindo atingir bitrates até 10 Mbits e elevados padrões de resolução espacial e temporal.

Os sistemas de codificação baseados em conteúdo permitem o tratamento individual de cada objecto, o que representa uma grande vantagem pois ao tratarmos de forma diferente objectos que necessitam de ferramentas diversas devido por exemplo às diferenças no seu movimento, podemos aumentar a eficiência de compressão.

Em seguida (Fig. II-11) é exibido o esquema geral de codificação no MPEG-4:

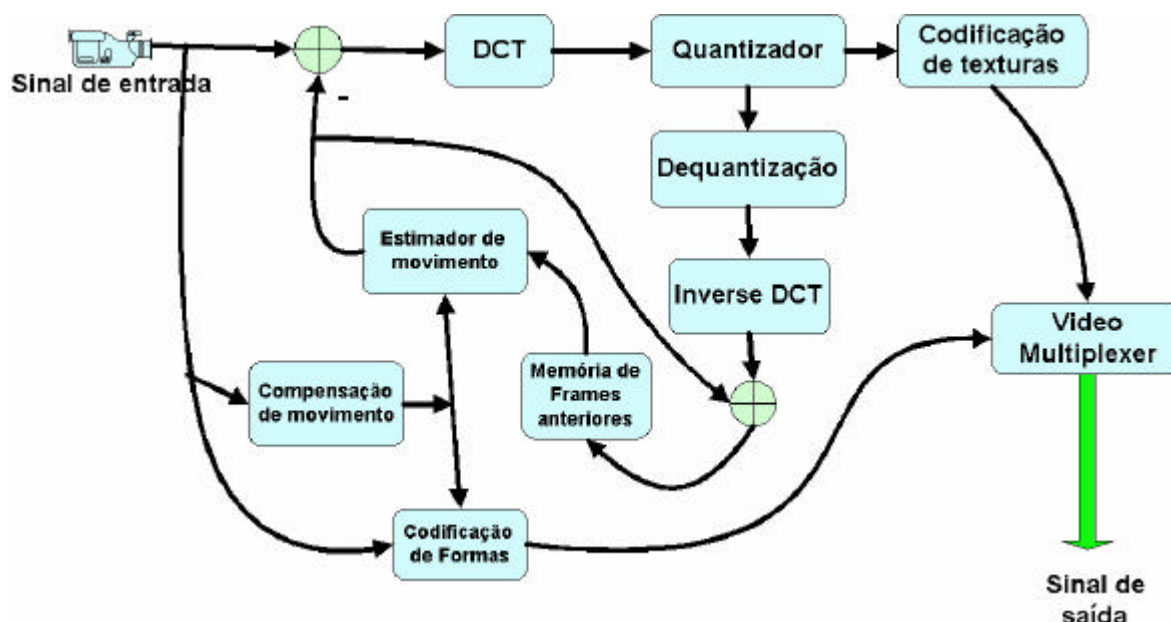


Fig II-11 – Codificador típico MPEG-4

Este esquema é a implementação prática da FGS tal como descrita no Capítulo III. No entanto, aqui existe codificação separada das formas e texturas de cada *Video Object* que são transmitidas de forma a permitir ao receptor o seu tratamento diferenciado.

Por exemplo, a norma define os *Static Sprites* como sendo partes da imagem que não sofrem alterações como por exemplo a paisagem de fundo de uma cena. Os *Static Sprites* podem ser tratados individualmente para serem transmitidos no primeiro quadro e depois serem alojados num buffer do receptor. Assim, é necessário transmitir muito pouca informação acerca deste objecto. Também cada objecto pode ser tratado via *Global Motion Compensation* (GMC) em que o movimento global do objecto é representado por um número reduzido de parâmetros. O receptor pode depois conjugar estes 2 tipos de objectos para ir formando as sucessivas imagens (Fig II-12):

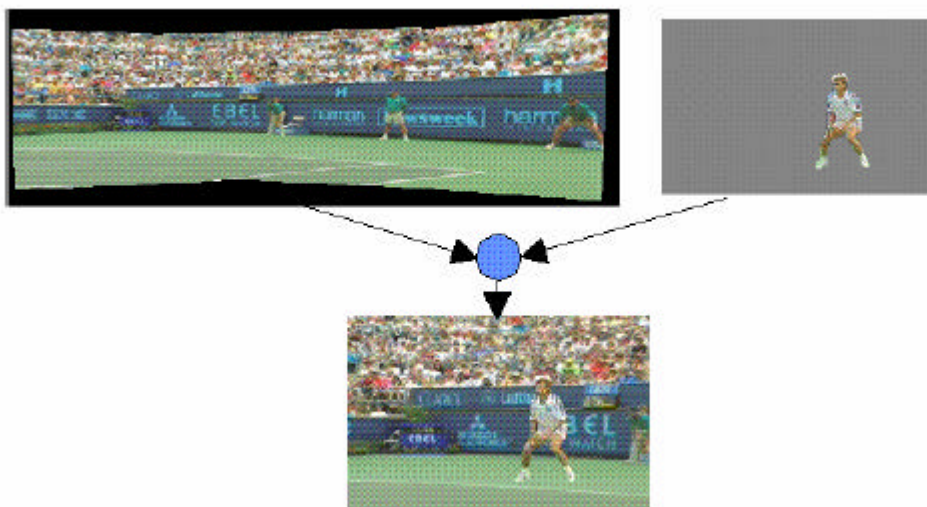


Fig. II-12 - Exemplo da aplicação de Codificação por Sprites [2]

Toda esta panóplia de inovações e possibilidades tornaram o MPEG-4 uma boa norma de *streaming* de vídeo na Internet. Para além da melhoria de eficiência nas técnicas de codificação e compressão, a sua adaptação a ritmos binários baixos onde ainda se encontram a maioria dos acessos à Internet e a sua capacidade para integrar os mecanismos de escalabilidade, revelam uma melhor adaptação às condições instáveis da rede.

Testes realizados pelo ISO [23, 24 e 25] revelaram desempenho superior às outras técnicas até então existentes no que diz respeito a robustez na resposta a erros, adaptação de escalabilidade e qualidade de imagem em baixos ou médios débitos.

Em seguida apresentamos uma evolução da norma MPEG-4, nomeadamente a sua parte 10 que permite melhorar o desempenho do MPEG-4 principalmente nos casos em que é possível tirar proveito das capacidades de máquinas com boas características técnicas.

II.4 Advanced Video Coding (H.264 ou MPEG-4 part 10)

O *Advanced Video Coding* (AVC), também designado por H.264 ou como a parte 10 do MPEG-4, é um novo formato de compressão de vídeo digital que resulta da união de esforços do denominado *Joint Video Team* que reuniu o *Motion Pictures Group*, a partir do desenvolvimento do MPEG-4, e do ITU/T e ISO na sequência dos anteriores desenvolvimentos do H.263.

O objectivo do AVC é permitir a transmissão de vídeo com qualidade DVD na Internet com a utilização de menores recursos [21]. É clara a intenção de cobrir o mercado de utilizadores de aplicações de vídeo em tempo real na Internet fixa ou nas redes sem fios.

Enquanto que o MPEG-4 tinha por objectivo cobrir os utilizadores com baixos débitos de ligação, o AVC pretende ir mais longe, não só para substituir o MPEG-4 Visual como também dar um novo salto tecnológico que permita obter melhores resultados que o MPEG-2 nas ligações com maiores recursos.

Tal como os seus antecessores, o AVC não define explicitamente um codec mas apenas a sintaxe da bitstream à saída do codificador e a sua forma de descodificação.

A implementação genérica é muito semelhante à dos seus antecessores, ou seja, predição, compensação de movimento, quantificação, redução de redundância, etc.. A evolução tecnológica reside em novas formas de implementar os vários elementos do processo de codificação e compressão tais como a DCT, a quantificação, a compensação de movimento ou a codificação. Por outro lado, a norma pretende ser completamente transparente em relação à rede de transporte para que tanto possa ser implementada em ambientes Internet (IP) ou em redes privadas como por exemplo o ATM.

Destas evoluções, e em relação ao MPEG-4, podemos indicar como mais importantes [18 e 20]:

- ? Compensação de movimento com tamanhos variáveis dos macroblocos: Desde o típico 16x16 até o mínimo de 4x4 com uma série de possibilidades distintas entre eles;
- ? Menor complexidade na implementação do processo de precisão dos movimentos até $\frac{1}{4}$ de pixel;
- ? A DCT é aplicada a blocos 4x4 e não a 8x8 como acontecia antes para uma codificação mais precisa; no entanto, são conferidos graus de liberdade para que possa ir até 16x16 na crominância ou 8x8 na luminância se tal for vantajoso;
- ? Utiliza coeficientes inteiros na DCT;

O AVC institui 2 camadas independentes [18]: a *Video Coding Layer* (VCL), responsável por uma eficiente representação do vídeo em si e a utilização da denominada NAL – *Network Abstraction Layer* que formata a *bitstream* produzida na codificação e lhe adiciona informação para adaptação à rede de transporte.

II.4.1 Codificação (VCL)

A primeira camada da norma AVC é a Vídeo Coding Layer, que engloba todo o processo de codificação de um vídeo a ser transmitido. Tal como nas anteriores versões, um vídeo é composto por imagens e objectos individuais que são divididos em blocos para facilitar o processo de compressão.

As inovações começam logo aqui: enquanto que no MPEG-2 ou no MPEG-4 os blocos para a compensação de movimento eram do tamanho 8x8, no AVC há flexibilidade na construção, pois permite a utilização de 8 tipos diferentes na predição por compensação de movimento (os blocos podem ser, por exemplo, 4x4, 4x8, 8x8 16x8 ou 16x16).

No que diz respeito aos vectores de movimento, há uma maior precisão, pois podem apontar para $\frac{1}{4}$ de pixel. O objectivo é tentar isolar e precisar o mais possível o movimento. No método de estimação, a referência não está sujeita à imagem anterior mas sim a múltiplas imagens que são armazenadas pelo codificador. Desta forma, pretende-se reduzir o erro associado à estimação de imagens e tratar de forma mais eficiente movimentos periódicos.

Ainda é possível utilizar estimação intraquadro, ou seja, a capacidade de representar uma zona de uma imagem a partir da informação vizinha nessa mesma imagem. Se existirem redundâncias espaciais, esse facto pode ser explorado na representação de blocos adjacentes iguais.

Outra evolução acontece na passagem do domínio espacial para o domínio da frequência. Utiliza uma transformada a partir de blocos 4x4 de coeficientes inteiros. Desta forma tenta-se eliminar os erros de arredondamento associado aos coeficientes não inteiros usados no MPEG-4 parte 2, por exemplo.

Para eliminar diferenças que possam surgir entre blocos que tenham sido construídos com diferentes tipos de estimação ou diferentes degraus de quantificação, existe o denominado *Loop Filter*. A sua função é a de limar as súbitas mudanças na imagem resultante ao actuar nos 2 pixels de cada lado da fronteira dos blocos.

Finalmente, foram estabelecidos 2 métodos para a construção da bitstream a partir dos coeficientes de quantificação, ambos utilizando variações dos códigos de comprimento variável [20]:

- ? CAVLC – Context Adaptative Variable Length Coding: Representa uma evolução do VLC utilizado no MPEG-4 Parte 2. Aqui, existe uma tabela de diferentes Códigos Variáveis. Esta tabela é construída para rentabilizar as relações estatísticas deste tipo de dados. A escolha do código da tabela a utilizar para uma determinada sequência é determinada pelos códigos já utilizados em sequências anteriores.
- ? CABAC – Context Adaptative Binary Arithmetic Coding: é um método ainda mais eficiente que o CAVLC (5 a 15 % na redução do bit rate). Para além da análise das probabilidades de ocorrência de símbolos tendo em conta os dados de sequências anteriores já codificadas, utiliza Códigos aritméticos o que permite palavras de código não inteiras no número de bits. Explora a correlação entre símbolos para diminuir o número de bits necessários à representação.

O resultado é um codificador com a seguinte estrutura básica (Fig II-13) [18]:

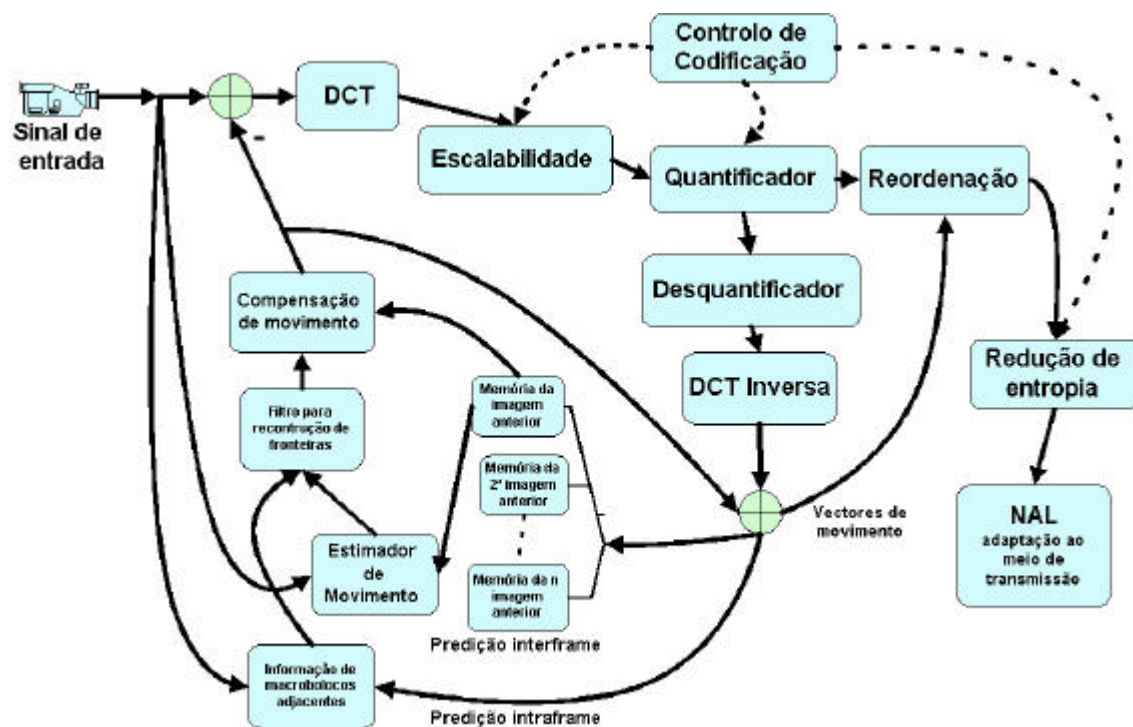


Fig II-13 - Codificador típico AVC

II.4.2 Adaptação ao meio de transmissão (NAL)

O MPEG-4 parte 10 define explicitamente uma camada cuja função é a de adaptação do bitstream gerado na codificação ao tipo de rede usada para a sua transmissão e que é designada por Network Abstraction Layer (NAL). Desta forma tenta-se blindar a construção da bitstream que assim se pode desenvolver sem os constrangimentos inerentes à necessidade de saber como irá ser transportada. Essa preocupação ficará a cargo de outra entidade, neste caso a NAL.

Os dados transmitidos são encapsulados em pacotes denominados *Unidades NAL* [18]. Cada um destes pacotes, constituído por um número inteiro de bytes, tem um cabeçalho de 1 byte cuja função é o de identificar perante a rede o tipo de dados que transporta. O restante do pacote, o payload, é o bitstream do vídeo codificado. No caso abordado por este trabalho, transmissão na Internet, cada Unidade NAL é transportada nos pacotes IP. Como estes têm por si só toda a identificação necessária relativa a origem, destino, tamanho do pacote e tipo de mensagem, a unidade NAL não necessita de especificar por exemplo o início e fim do seu payload.

Existem 2 tipos de Unidades NAL: As unidades VCL NAL, que transportam o bitstream do vídeo codificado, e as não VCL NAL que transportam informações adicionais para os decodificadores, tais como informações temporais ou outras que melhorem o desempenho da decodificação.

Um conjunto de Unidades NAL que constitui uma imagem é denominado por uma Unidade de Acesso. Esta pode incluir elementos antes das unidades VCL NAL para identificar o seu tamanho ou referencias temporais dessa imagem. Por sua vez, uma sequência de Unidades de Acesso constrói a sequência de vídeo codificado. Cada uma destas representa quantidades de vídeo que o decodificador pode interpretar independentemente.

Numa parte do quarto Capítulo analisaremos a implementação do AVC na transmissão de vídeo em tempo real na Internet e as possibilidades que oferece aos utilizadores.

II.5 – Codecs Proprietários

Enquanto o MPEG-4 é uma norma aberta, não proprietária e que resulta do esforço de várias entidades, as 3 empresas responsáveis pelos principais codecs proprietários a seguir descritos (QuickTime, Real Networks e Windows Media Player) estão em constante competição para apresentarem soluções cada vez mais eficientes e robustas para lhes permitir aceder a uma maior fatia do mercado de *streaming*. Nos 3 casos foram desenvolvidas ferramentas completas para *streaming* desde o codificador, a forma de transmissão e o decodificador. Devido aos sistemas serem proprietários não nos permite detalhar as normas de compressão ou os constituintes dos codificadores ou decodificadores. De seguida, serão apresentadas, sumariamente, as características gerais destes formatos.

II.5.1 QuickTime

É uma solução desenvolvida pela Apple desde 1991 como uma plataforma chave na mão para criação, visualização e *streaming* de conteúdos de média em redes IP, wireless ou de banda larga. Também é considerada uma plataforma aberta dado que disponibiliza o

código fonte de alguns componentes e tem dado grande ajuda ao desenvolvimento da norma MPEG-4 que incorpora na sua estrutura na versão QuickTime 6 procurando satisfazer os utilizadores de Internet de baixos recursos. A Apple dá grande relevo à total integração da norma MPEG 4 incluindo as facilidades de escalabilidade, protecção de erros e compressão orientada a objectos, por exemplo. No entanto, introduziu algumas melhorias, como por exemplo:

- ? Gestão mais rigorosa de cores;
- ? Estimação de movimento mais precisa e rápida;
- ? Quantificação melhorada;
- ? Suporte de AAC (Advanced Audio Coding ou MPEG-2 Áudio);

Dos 3 formatos é o que utiliza o *buffer* mais curto, beneficiando a rápida inicialização da visualização [15]. Outro desenvolvimento feito pela Apple é o denominado Skip Protection, um mecanismo que pretende reduzir ao mínimo as interrupções na apresentação devidas a congestionamento da rede. Isto acontece porque o codificador aproveita os instantes em que a largura de banda disponível é superior à taxa de leitura do decodificador para lhe enviar mais informação que fica residente num buffer. Quando a rede congestionada, o decodificador recorre a esta informação para não parar.

Nesta última versão estão incluídos o editor de vídeo para *streaming* QuickTime Pro, o servidor QuickTime *Streaming* Server 4 e o QuickTime Broadcaster para eventos ao vivo incorporando já a possibilidade de utilizar a linguagem SMIL que é um standard em franca expansão para edição de páginas multimédia.

II.5.2 RealNetworks

A plataforma desenvolvida pela Real Networks está largamente difundida, calculando-se que 90% dos PC's com leitores de *streaming* tenham instalado o decodificador Real Player [11]. A Real Networks está essencialmente dedicada ao desenvolvimento de produtos para tecnologia de *streaming* de média e gera os seus lucros a partir da comercialização de leitores e codificadores para *streaming*.

O primeiro Real Player data de 1995. A última versão do leitor é o Real Vídeo 10 cujas principais características são [13]:

- ? É um codificador com compensação de movimento;
- ? Métodos de predição espacial proprietários;
- ? Codificação com redução de entropia adaptada a cada caso;
- ? Esquemas de filtragem para adaptação às características psico-visuais humanas;
- ? Codificação a 2 passos;
- ? Suporte de vídeo entrelaçado.

O codificador suporta taxas de transmissão constantes (CBR), ou variáveis (VBR) e codificação baseada na qualidade. Permite a codificação a 2 passos, ou seja, o codificador realiza duas operações de codificação: na primeira realiza uma análise ao conteúdo e na segunda comprime-o com base nessa informação prévia para atingir melhores resultados. Este método aumenta o tempo de codificação e não é aplicável a eventos em tempo real.

O Codificador Real Helix Universal Server utiliza tecnologia escalável incluindo o formato MPEG-4 e permite também facilmente criar várias versões no servidor adaptadas a diferentes condições de transmissão e com a possibilidade de o utilizador final poder passar por diferentes representações ao longo do mesmo vídeo. Foi disponibilizada pela Real Networks, em formato *open source*, uma versão limitada deste codificador para permitir desenvolvimentos de outras entidades.

Do lado do codificador, foi desenvolvida a capacidade de estimar a largura de banda de uma determinada ligação. Desta forma é mais fácil evitar a colocação na rede de dados que depois não venham a ser utilizados. No que diz respeito à transmissão, o seu ritmo consegue adaptar-se às variações de velocidade inerentes à Internet e, assim, consegue obter razoáveis desempenhos em ligações de baixo débito.

No que diz respeito ao formato da imagem, o Real Video 10 é capaz de suportar vídeos com dimensões interessantes para ligações do tipo Dial. Por exemplo, para uma taxa de transmissão de 48 Kbps, é possível visualizar uma imagem de baixo movimento com a dimensão CIF a 30 imagens por segundo. O topo da qualidade é a televisão digital de alta definição: a Real Networks afirma que a sua técnica de compressão é capaz de armazenar mais de 15 filmes completos num único DVD. [13]

Entre as desvantagens do formato, podemos indicar que a compressão é iniciada pelo áudio o que pode deixar em desvantagem o vídeo caso o espaço disponível ou a largura de banda sejam muito baixos. O controlo da largura de banda da transmissão feita a partir do codificador tem o inconveniente de sobrecarregar a rede com mensagens de controlo. Do lado do leitor é utilizado um buffer bastante grande (cerca de 15 segundos) o que atrasa o início da visualização.

A Real Networks afirma [13 e 14] que os resultados de testes demonstram que o Real Video 10 atinge os seguintes desempenhos:

- ? Redução de 30 % no tamanho do bitstream para a mesma qualidade do Windows Media Video 9;
- ? Redução de 80 % no tamanho do bitstream para a mesma qualidade do MPEG 2;
- ? Redução de 45 % no tamanho do bitstream para a mesma qualidade do MPEG 4;
- ? Redução de 15 % no tamanho do bitstream para a mesma qualidade do H.264.

II.5.3 Windows Media

A plataforma Windows Media da Microsoft é um conjunto de ferramentas para a edição e transmissão de vídeo que beneficia do facto de estar associada aos produtos desta empresa, nomeadamente os sistemas operativos Windows, o que lhe garante uma enorme penetração no mercado de leitores de *streaming*.

Através dos seus protocolos proprietários, a Microsoft pretende criar uma ferramenta universal para todos os passos do processo de *streaming* (criação, edição, transporte e apresentação) que seja capaz de produzir melhores resultados que o MPEG-4.

A última versão, o Windows Media 9, inclui alguns desenvolvimentos que lhe permitem suportar múltiplas velocidades de transporte numa única sequência de dados.

No que diz respeito à criação e edição de vídeo, o Windows Media Video 9 (WMV9) é o mais popular dos codecs disponibilizados pela Microsoft. A sua versatilidade fica patente na gama de possibilidades para o formato da imagem final, que varia entre [10]:

- ? 160x120 pixels para velocidades de transmissão de 10 Kbps
- ? 1920x1080 pixels para velocidades até 20 Mbits

O codificador inclui ferramentas proprietárias para codificação de vídeo no modo entrelaçado no modo 4:1:1. Tal como no MPEG, também a Microsoft definiu níveis de qualidade e complexidade da codificação:

- ? Nível Baixo – suporta QCIF a 96 Kbps e 15 imagens/segundo
- ? Nível Principal – desempenho equivalente à norma MPEG-2
- ? Nível Alto – apropriado para aplicações de alta definição.

A possibilidade de codificar vídeos em Variable Bit Rate (VBR) não é nova. No entanto, a Microsoft introduz 3 variantes deste método:

- ? VBR baseado em qualidade : permite garantir a mesma qualidade ao longo de uma sessão, aumentando se necessário a taxa de transmissão;
- ? VBR baseado em taxa de transmissão (2 passos) : sistema que tenta atingir permanentemente a melhor qualidade em função da taxa de transmissão disponível no momento (útil em algumas situações de transmissão na Internet mas não em transmissões ao vivo);
- ? VBR baseado em picos de velocidade (2 passos) : igual ao anterior mas em que fica definida a máxima velocidade possível (mais aplicada, por exemplo, em leitores de CD em que existe uma taxa máxima de leitura).

O WMV9 suporta conteúdos com a tecnologia de vídeo entrelaçado e pixels não quadrados, o que permite que a captura de imagens feitas no formato *Digital Video* (DV) sofra menos distorção [12].

Uma inovação apregoada pela Microsoft, e que poderá ser importante na transmissão via Internet, é o denominado “Video Smoothing” que é a capacidade de o decodificador poder reconstruir imagens perdidas na transmissão através de um processo complexo de estimação com base na informação pixel-a-pixel das imagens vizinhas. Este processo pode, inclusive, ser usado para aumentar a qualidade de uma imagem codificada com baixas taxas de imagens por segundo, pois no fundo é uma técnica de interpolar imagens construídas pelo decodificador. Contudo, este processo requer muita capacidade do CPU do PC, exigindo no mínimo um processador a 733 MHz.

Esta oferta integrada, inclui leitores e codificadores de fácil utilização e bom desempenho, especialmente a componente áudio que já permite o sistema 7.1 *surround sound*. Também aqui, o buffer do leitor Windows MediaPlayer tem um tamanho que atrasa o início da apresentação da sessão.

Foi dada relevância à protecção dos direitos de autor pois a ferramenta *Digital Rights Management* inclui um conjunto completo de utensílios para implementar o aluguer ou venda de conteúdos [12].

A Microsoft defende que o Windows Media 9 é três vezes melhor que o MPEG-2 e duas vezes superior ao MPEG-4, principalmente para as taxas de transmissão mais elevadas [10].

Capítulo III - *Streaming* de Vídeo

Quando se iniciou a implementação de mecanismos de transmissão de vídeo na Internet, as imagens captadas por meios convencionais eram digitalizadas e, após a sua codificação, eram guardadas em forma de ficheiros em máquinas servidoras. Estes ficheiros eram descarregados para a máquina cliente após um pedido desta ao servidor e o utilizador era obrigado a aguardar pela descarga completa do ficheiro para que depois um software adequado na sua máquina fizesse a descodificação e a sua leitura para a apresentação.

Este método tem vários inconvenientes:

- ? São necessários grandes espaços de armazenamento nos servidores tanto do lado do cliente como do lado do servidor, pois o tamanho dos ficheiros de vídeo, mesmo comprimidos, é sempre considerável;
- ? O tempo que o utilizador tem de esperar desde que solicita o vídeo até começar a sua visualização é, em regra, enorme, pois tem de se aguardar até à chegada de um ficheiro inteiro transmitido por uma rede de comportamento variável e com velocidades limitadas principalmente no troço de cliente;
- ? Não permite a transmissão de eventos em tempo real pois requer que a totalidade do ficheiro a ser enviado ao cliente já esteja disponível no servidor aquando da solicitação deste.

Desenvolvimentos tecnológicos permitiram a evolução no sentido da disponibilização de conteúdos de vídeo na Internet sob a forma de *streaming* vídeo, uma tecnologia surgida no início dos anos 90[1] que permite a transmissão de vídeo em tempo real e a sua apresentação *on-the-fly*, isto é, a possibilidade de se iniciar a apresentação de um objecto enquanto decorre ainda a sua transmissão. Neste tipo de tecnologia, o receptor solicita a transmissão de um determinado vídeo e o servidor inicia a sua transmissão partindo o ficheiro em pacotes que são agrupados em trechos (*streams*), ou seja, pedaços que podem ser apresentados individualmente independentemente de se ter já visualizado os trechos anteriores. A aplicação terá de fazer a gestão dos pacotes recebidos, reconstruindo os trechos e colocando-os por ordem de apresentação.

Contudo, a Internet continua a colocar grandes obstáculos a esta tecnologia, principalmente porque é uma rede do tipo *best-effort* que não pode garantir uma largura de banda estável nem tempos de resposta diminutos e constantes, ou seja, não implementa mecanismos de Qualidade de Serviço. Se não forem utilizadas ferramentas para ultrapassar estas condições, tais como, protocolos adaptados, servidores melhorados ou mecanismos de compressão e codificação, a transmissão de vídeo em tempo real na Internet seria uma aventura com resultados imprevisíveis.

Actualmente, esta tecnologia está em franca expansão no número cada vez maior de aderentes e no volume de tráfego gerado na rede. Para isto também contribuiu a evolução tecnológica dos PC's, que passaram a possuir mais capacidade de processamento, memória e de disco, e também, o desenvolvimento da Internet que permitiu difundir por todo o mundo estas aplicações. Por exemplo, nos nossos dias, praticamente todos os PC's chegam ao utilizador com software para leitura de vídeo *streaming*, ou é muito fácil a sua obtenção por download via web.

O *streaming* está perfeitamente implantado como a solução para a transmissão de vídeo em tempo real na Internet por ser a tecnologia que, por enquanto, melhor responde às necessidades dos utilizadores e fornecedores de conteúdos de média.

III.1 - Funcionamento do streaming

O *streaming* de vídeo é feito segundo o esquema da figura III.1:

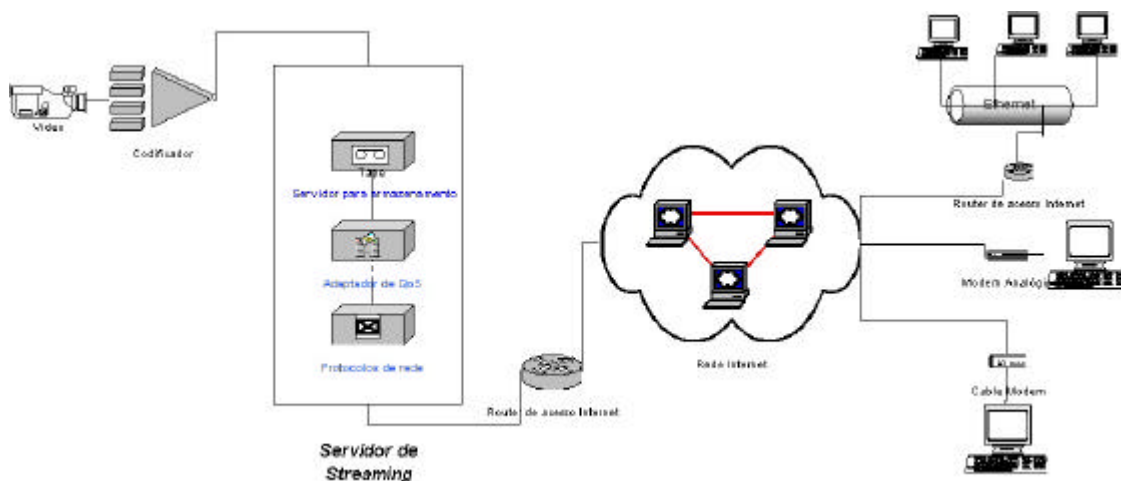


Figura III-1 – Esquema geral de streaming de vídeo

Podemos, sucintamente, descrever o mecanismo representado da seguinte forma [2]: as imagens de vídeo e os sons capturados por câmaras de vídeo digitais são comprimidos e colocados em servidores. A pedido do cliente final, o servidor de *streaming* retira a informação deste local para a entregar a uma aplicação cuja função é adaptar a taxa de transmissão às condições momentâneas do canal de transmissão. Os protocolos empacotam a informação para o seu transporte pela rede IP. Nesta rede, em locais estratégicos, são colocadas máquinas capazes de fazer *cache* para jusante, o que melhora o desempenho da rede: são os denominados serviços de distribuição contínuos. Finalmente, já na máquina receptora, a aplicação de leitura de *streaming* recebe as streams, ordena-as, sincroniza o vídeo e o áudio e entrega-o ao decodificador (*decoder*) para apresentação.

As diferentes componentes de um sistema de distribuição de *streaming* são:

- ? Compressão de vídeo
- ? Servidores de *streaming*
- ? Adaptação ao meio de transmissão
- ? Serviços de distribuição contínuos
- ? Mecanismos de sincronização
- ? Protocolos

Há parâmetros no sistema cujo desempenho pode determinar o sucesso ou insucesso de uma boa transmissão de *streaming* de vídeo:

- ? *Largura de Banda*: para se conseguir uma qualidade mínima e uma transmissão perceptível para o receptor humano, há exigência de um mínimo de largura de banda que a rede Internet não consegue garantir. Há mesmo a necessidade de implementar mecanismos de controlo de congestionamento nos períodos mais sobrecarregados de tráfego. Nos mecanismos de *streaming* este controlo é feito limitando a taxa de transmissão.
- ? *Atrasos* : Este é um factor determinante pois, ao contrário de outras aplicações de dados, um pacote que não chegue a tempo de ser decodificado e apresentado, cria uma descontinuidade não recuperável. Como a rede Internet induz atrasos inconstantes, há a necessidade de se introduzir no receptor um *buffer*.
- ? *Perdas de pacotes*: As perdas de pacotes são inerentes à Internet o que implica que o protocolo de *streaming* tenha de responder às perdas de forma robusta.
- ? *Funções típicas de um VCR*: os utilizadores de vídeo habituaram-se, nos seus aparelhos de vídeo caseiros, às funções de *rewind*, *fastforward* ou acesso a determinados trechos. Para que sejam implementáveis em *streaming*, também foram definidos alguns formatos.
- ? *Complexidade na decodificação*: para que estas aplicações possam ser implementadas em diferentes tipos de aparelhos (PC's, PDA's, Telemóveis) a decodificação deverá ser o mais simples possível e consumir o mínimo de recursos.

Cada um destes aspectos desempenha um papel fundamental e actualmente estão em constante desenvolvimento tendo em vista a maximização do seu desempenho.

III.1.1 Codificação de Vídeo adaptada a *streaming*

Nas técnicas de *streaming*, há dois grandes tipos de compressão [2]: escalável e não escalável. A compressão não escalável origina apenas uma representação do objecto, o que inviabiliza qualquer tipo de adaptação às condições inerentes ao utilizador final. A compressão é feita para uma determinada taxa o que origina dois efeitos:

- ? Apenas permite a visualização do vídeo por utilizadores capazes de tratar a taxa de transmissão resultante;
- ? Não permite aos clientes que usufruam de ligações de melhor qualidade tirar partido disso.

Por isso, surgiu a compressão escalável, onde a codificação transforma as imagens vídeo em múltiplas subrepresentações de uma representação base. Esta representação base pode por si só servir para o cliente ver as imagens num leitor apropriado e é o mínimo que o receptor precisa para poder ver uma representação do vídeo pedido. Cada uma das subrepresentações servirá para melhorar a qualidade e a complexidade das imagens reproduzidas em conjunto com a base.

A ideia subjacente é a de apenas enviar para o cliente as subrepresentações que o canal de transmissão e o equipamento suportarem para assim termos uma adaptação às condições individuais, quer de largura de banda quer de perdas. Daqui resulta que, embora continue a existir um patamar mínimo ao qual as condições de um cliente particular têm que satisfazer, pode-se começar a proporcionar às melhores ligações um resultado final de melhor qualidade.

Na seguinte figura (Fig III-2) ilustramos graficamente o desempenho desejado de um sistema de *streaming* de vídeo e o resultado obtido com técnicas de compressão não escaláveis e escaláveis:

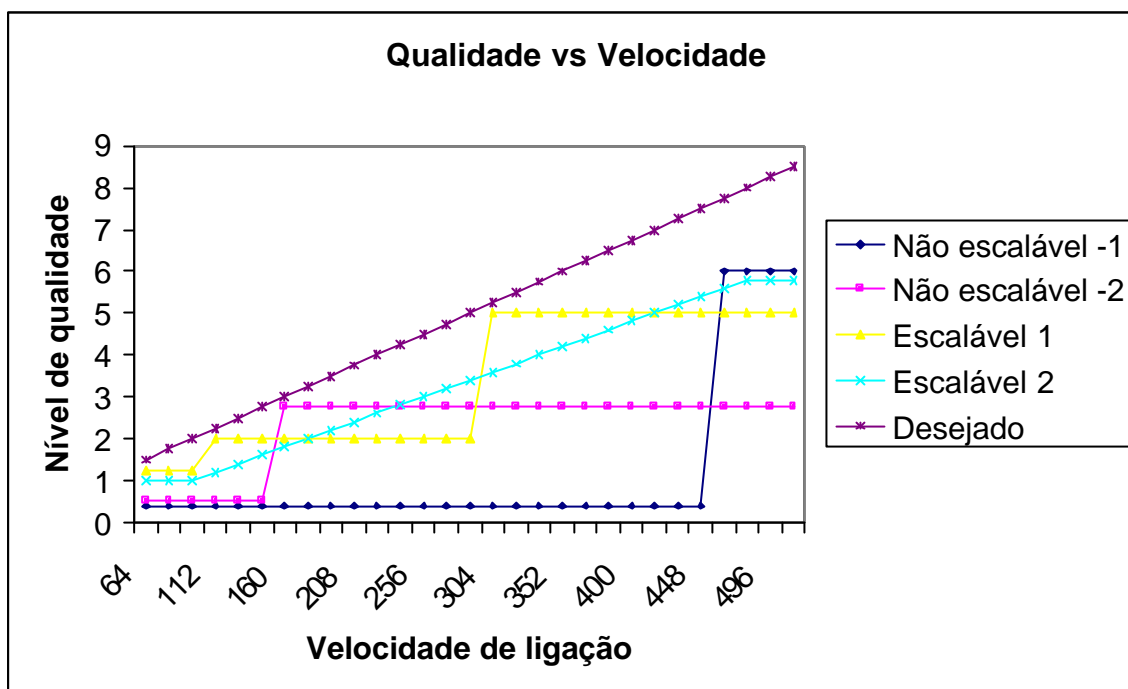


Fig. III-2 – Qualidade/Velocidade para compressões escaláveis e não escaláveis

Com as técnicas não escaláveis existe um ponto de *cutoff*, abaixo do qual não podemos ver o vídeo desejado, e acima do qual é possível, mas sem alteração da qualidade do mesmo. Se este ponto estiver localizado nas baixas capacidades de ligação, permitimos a visualização do vídeo a um maior número de clientes mas com baixa qualidade de apresentação. Pelo contrário, se privilegiarmos a qualidade de apresentação impedimos a maior parte dos interessados de a ele aceder.

Pelo contrário, as técnicas escaláveis tendem a adaptar-se ao canal de transmissão dando menos qualidade às piores ligações e fornecendo um vídeo de qualidade para quem puder suportar maiores volumes de informação. Na representação final, esta escalabilidade pode manifestar-se na variação da qualidade da imagem, no tamanho da mesma ou no número de imagens por segundo.

A forma de compressão das imagens para posterior codificação e transmissão via Internet tem de obedecer a algumas exigências quando comparadas com os algoritmos utilizados pela transmissão de televisão tais como uma grande escalabilidade, leveza para tratamento pelos PC's e capacidade de ultrapassar as inevitáveis perdas de pacotes e atrasos da rede. Os standards existentes até há pouco, H.261, H.263 e as várias evoluções do MPEG, tiveram de se adaptar e evoluir para responder a estes requisitos

Destes, o MPEG tem tido uma evolução no sentido de se converter num protocolo de compressão adaptado à Internet. O MPEG-4 foi mesmo pensado para possibilitar videoconferência em ligações de 64 Kbps para se tornar o protocolo por excelência de

compressão de vídeo na Internet a médio prazo. Uma explicação técnica já foi feita com maior detalhe no Capítulo II deste trabalho.

III.1.2 Codecs de Vídeo para *Streaming* de Média

O codec é o equipamento responsável pela codificação e/ou decodificação do sinal de vídeo para que ele seja transportado pelo canal de transmissão com a melhor qualidade possível. No lado emissor, o codec tem de transformar o sinal de vídeo recebido pela fonte, digitalizá-lo caso a fonte seja analógica, codificar o sinal segundo formatos estabelecidos e posteriormente adaptar o seu envio às características do canal.

No caso da Internet, os codecs têm de resolver ou minimizar as dificuldades já descritas em capítulos anteriores. Por outro lado, a própria natureza das imagens origina diferentes quantidades de informação a transmitir durante a mesma sessão.

Como já vimos atrás, a implementação do *preroll buffer* à entrada do decodificador atenua o problema da taxa de transmissão descontínua da Internet. Desta forma, o emissor apenas tem que produzir informação a taxa constante.

Posteriormente, colocam-se problemas relacionados com transmissões em tempo real: como permitir a ligação de clientes distintos em tempos distintos ou como permitir as características semelhantes a um VCR tais como o *rewind*, o *fastforward*, etc.?

III.1.3 Servidores para *Streaming*

O surgimento e desenvolvimento de aplicações orientadas para o *streaming* de media apenas ocorreu em 1995. Só nesta altura se passou a fazer desenvolvimento conjunto para uma tecnologia desde o servidor ao PC cliente.

A necessidade de enviar altos débitos de informação para satisfazer pedidos de vários clientes que podem aceder em tempos diferentes leva tipicamente à utilização de servidores dedicados para o *streaming* com altas capacidades de processamento e possuidores de elevados recursos de armazenamento. Esta é uma importante peça da construção do processo de *streaming* pois o processamento de dados multimédia e a sua disponibilização à rede de uma forma rápida e eficaz representa um aspecto fundamental para o sucesso do *streaming*.

Podemos desmontar a arquitectura de um servidor de *streaming* em três componentes [2]:

- ? Comunicação
- ? Sistema Operativo
- ? Sistema de armazenamento

Comunicação

Cabem aqui todas as funções de aplicação, ou seja, a forma de representar as imagens após terem sido capturadas e também todas as formas de comunicação com o utilizador, nomeadamente protocolos, sincronização áudio e vídeo, forma de troca de informação com o utilizador, etc.

Sistema Operativo

Um sistema operativo, base para a construção de todo o tipo de serviços num computador, tem também de gerir e comandar os recursos que lhe estão afectos. Para além das funções

típicas de um sistema operativo, em *streaming* terá também de possibilitar o acesso e o recurso às aplicações em tempo real.

O estudo do sistema operativo de um servidor de *streaming* tem três capítulos: o controlo de processos, a gestão de recursos e a gestão de ficheiros.

No primeiro, o controlo da CPU tem como objectivo primordial a gestão de tarefas por nível de prioridade em tempo real. Para isso são utilizadas as técnicas de EDF (Earliest Deadline First) onde a tarefa que tem de ser terminada primeiro é a prioritária ou Rate Monotonic Scheduling em que a tarefa mais vezes solicitada é aquela que recebe a preferência temporal.

A gestão de recursos envolve, como o nome indica, a gestão da CPU, das memórias e dos discos presentes no servidor. Em *streaming*, esta gestão traduz-se pelo controlo de número de clientes ligados para evitar degradação de QoS e a gestão de recursos alocados ao processo. Estes aspectos estão interligados pois cada vez que há um novo pedido de ligação, o servidor terá de fazer um teste para saber se pode garantir a qualidade de serviço que é exigida na transmissão e em caso de o aceitar, terá de reservar no sistema os recursos necessários a esse cliente. A admissão de novos utilizadores pode ser determinística, ou seja por um número total pré-estabelecido, ou estatisticamente tendo em conta um compromisso entre a qualidade e o número de utilizadores permitidos.

Sistema de armazenamento

Para corresponder às necessidades do *streaming* de vídeo, a componente de armazenamento num servidor é concebida para fornecer a informação a uma elevada velocidade simultaneamente a vários receptores, para ter uma larga capacidade e ser tolerante a falhas próprias.

Cada um dos ficheiros de áudio ou vídeo pode ser guardado num único disco, o que obriga a existência de correspondência entre o ficheiro e o respectivo disco, ou então, cada ficheiro é repartido por vários discos por forma a reduzir o tempo de pesquisa dos mesmos através de distribuição de tarefas. Actualmente, a forma de garantir um rápido desempenho de um servidor perante múltiplas solicitações passa por uma técnica designada *Data Striping* e que consiste em dividir a informação relativa a um ficheiro por vários discos que podem ser acedidos paralelamente. A gestão de ficheiros garante ao utilizador que não receberá um trecho de vídeo diferente daquele que pediu. A gestão dos discos e do armazenamento dos ficheiros é feita para privilegiar o funcionalismo das aplicações dado que temos um sistema em tempo real.

Para aumentar as capacidades de armazenamento, poderiam colocar-se vários discos, o que, no entanto, provoca um disparar de custos que é geralmente incomportável. A solução de maximização por custo reduzido recorre à implementação de um sistema em que apenas uma pequena parte da informação está guardada no(s) disco(s) do servidor enquanto que a restante se encontra em sistemas denominados terciários tais como tapes ou sistemas de *jukebox* de CD's. A operacionalização entre estes sistemas tem de ser optimizada estando mesmo indicadas formas de implementar SAN's, (*Storage Area Networks*) com elevados débitos entre os dispositivos de armazenamento ligados entre si por fibra óptica. Foi definido para este caso um protocolo específico de ligação, o NAS's (*Network Attached Storage*) para fazer comunicar os vários dispositivos que estão distribuídos por uma LAN ou mesmo por uma WAN, ligadas por Ethernet, ATM ou FDDI suportado em protocolos standard como HTTP ou TCP/UDP.

Por último, e para evitar descontinuidade de serviço prestado, o sistema de armazenamento do servidor deverá tornar eventuais falhas tais como erros de discos. Mais uma vez, a duplicação de discos parece a solução óbvia para implementar, mas tem custos acrescidos o que leva, por vezes, a utilizar técnicas de correcção de erros como o FEC (*Forward Error Control*) que implica a utilização de bits de paridade, ou seja um maior *overhead* no volume de tráfego tal como será descrito no ponto III.3.2.

III.1.4 Escalabilidade e Adaptação da taxa de transmissão

Cada vez mais são diversificados os meios de acesso à Internet: acesso via GSM a 9600 bps, ligações comutadas analógicas até 56 Kbps, ligações RDIS a 128 Kbps, ligações dedicadas *Frame Relay* ou Circuito Dedicado até 2 Mbps, ADSL até 1 Mbps ou ligações de *Broadband* até 155 Mbps. Por outro lado, cada um destes tipos de acesso pode ser variante no tempo em função da situação dos servidores, backbones, etc.. Ou seja, pode-se afirmar que cada cliente de um servidor de *streaming* é um caso único e o ideal é que a transmissão seja diferenciada e escalável.

A primeira solução pensada foi a criação de diferentes versões do mesmo vídeo no servidor cada uma delas adaptada a cada tipo de ligação. Ora este tipo de abordagem partia do pressuposto que a ligação era limitada pelo último troço, aquele que o cliente utiliza para se ligar ao seu *Internet Service Provider*, o que nem sempre é verdade. Por outro lado, não previa alterações na largura de banda disponível ou na percentagem de perda de pacotes durante a transmissão.

A evolução natural nesta área foram os sistemas realimentados em que a informação das condições de transmissão eram recolhidas de forma a adaptar a taxa de transmissão. Um exemplo desta técnica é o Sure Stream desenvolvido pela RealSystem. Este sistema produz múltiplas representações do objecto para responder a diferentes condições. Durante a sessão, o cliente monitoriza a largura de banda e a perda de pacotes e dá instruções ao servidor para lhe enviar a informação mais adaptada.

III.1.5 Mecanismos de distribuição de *streaming* multimédia

A distribuição de eventos de vídeo na Internet pode ser de dois tipos[1]:

- ? *Live Broadcasting*
- ? *On-demand Streaming*

Live Broadcasting

No *live broadcasting*, um evento é transmitido em tempo real. Depois de capturado pela fonte de vídeo, o codec encarrega-se de digitalizar, comprimir ou codificar este sinal para entregar ao servidor de *streaming*. Este servidor inicial pode retransmitir a informação até ao utilizador final, mas geralmente o que ele faz é a entrega a servidores intermédios tal como vimos descrito no ponto anterior. No primeiro caso, os parâmetros de transmissão podem ser avaliados e registados pelo servidor. No segundo, o servidor inicial pode até nem saber quantos clientes finais receberam a informação enviada e muito menos as características de qualidade subjacentes a cada uma das entregas.

Foram desenvolvidas algumas técnicas relacionadas com a distribuição de eventos em tempo real:

- ? *Simulcast* : trata-se da criação de diferentes versões da informação adaptadas a diferentes velocidades de transmissão;
- ? *Push-splitting* : quando uma transmissão de uma amostra do evento é despoletada pelo servidor;
- ? *Pull – splitting* : quando o envio da informação é despoletada pelo cliente final;

On – demand Streaming

Ao contrário do live broadcasting, numa situação de on-demand a informação que é enviada pelo servidor foi previamente registada e guardada no servidor. Cada cliente faz um pedido ao servidor indicando qual o ficheiro que pretende receber. A ligação entre o servidor e o codec não existe. Isto permite que os servidores intermédios, actuando de *proxy servers*, façam cache da informação. As funções geralmente comuns em VCR, como o *fastforward* ou o *rewind* são desta forma possíveis de implementar. Esta transmissão é sempre iniciada pelo cliente.

III.1.6 Sincronização de média

Em *streaming* de vídeo, as componentes de áudio e imagem são guardadas e transmitidas separadamente o que leva à implementação de mecanismos que garantam a sua sincronização na apresentação final, ou seja, que a relação temporal entre som e imagem seja mantida desde a gravação até à apresentação. Embora o sincronismo tenha de ser implementado em última instância pelo leitor final do *streaming*, todo o processo começa pela definição temporal entre as várias componentes do objecto. Esta operação é feita pelo aparelho gravador das imagens que pode posteriormente sofrer intervenção humana para determinadas alterações como introdução de slides ou de caracteres na imagem.

Num sentido mais abrangente, pode-se falar de 3 tipos de sincronismo [2]:

- ? Sincronismos intra-trecho: responsável pela apresentação no tempo correcto das várias frames na apresentação.
- ? Sincronismo inter-trecho: responsável pelo sincronismo entre os diferentes trechos que constituem o todo. Por exemplo, sincronismo entre a imagem e o som.
- ? Sincronização inter-objectos: necessária quando existem, para além de trechos dependentes do tempo como o vídeo e o áudio, objectos estáticos na apresentação como textos que não podem ultrapassar uma determinada tolerância temporal.

No que diz respeito ao *streaming* via Internet, o problema do sincronismo é enfatizado pela natureza da rede que é best effort, introduzindo tempos de atraso imprevisíveis e muito variantes. Por isso, é importante que todos os componentes da rede suportem as relações de sincronismo contribuindo para a redução de atrasos. Contudo, nos dias de hoje, a Internet ainda não é capaz de por si só evitar estes problemas pelo que quase toda a inteligência do sincronismo está nos sistemas finais de forma preventiva ou correctiva.

Os mecanismos preventivos podem ser distribuídos ao longo de toda a cadeia de transmissão desde os servidores até aos receptores para prever ou evitar atrasos na transmissão.

Há também os denominados mecanismos correctivos em que se tenta recuperar a informação inicial na presença de erros de sincronismo. Para isto foi desenvolvido um

protocolo específico: o SSP (Stream Synchronization Protocol) onde são induzidos atrasos específicos em determinadas trechos tendo em conta a observação das condições da rede no receptor. Por outras palavras, quando na recepção se detecta um atraso de um determinado trecho (p.e. vídeo) é introduzido um atraso artificial no som por forma a esperar pela componente atrasada.

Dado o crescimento exponencial da utilização das técnicas de *streaming* de vídeo na Internet, os investigadores desta tecnologia têm vindo constantemente a desenvolver soluções para os problemas que identificámos nos Capítulos anteriores. Nos próximos pontos deste Capítulo serão apresentados alguns desses desenvolvimentos específicos para o *streaming* de vídeo na Internet e que contribuíram para o seu aperfeiçoamento.

III.2 Desenvolvimentos na compressão de vídeo

Já foi atrás referido que é ao nível das técnicas de compressão de vídeo que reside o grande esforço para melhorar a transmissão de vídeo em tempo real na Internet. As técnicas de compressão não escalável são desaconselhadas pois não permitem adaptar-se às condições de ligação dos clientes e por isso apenas serão aqui tratadas as técnicas de compressão escalável.

Para qualquer tipo de compressão escalável de um ficheiro de vídeo, este é comprimido pelo codificador dando origem a pelo menos 2 subrepresentações, uma das quais, a base, tem de ser totalmente recebida pelo decodificador:

1. **Camada de base:** representa a qualidade mínima para a apresentação do vídeo no receptor. Esta camada, que tem de ser recebida totalmente, permite ao cliente visualizar o vídeo, com uma qualidade mínima aceitável;
2. **Camada de melhoria:** Permite acrescentar qualidade à representação da camada de base mas não é prioritária nem mandatária para a apresentação do objecto de vídeo como acontece com a camada de base. Os seja, se as condições de transmissão permitirem que seja recebida em tempo útil, ela trará importantes benefícios à qualidade da apresentação do vídeo no aparelho receptor.

O desenvolvimento das técnicas de compressão escalável baseia-se na forma como estas camadas são originadas do lado do emissor e como são tratadas pelo receptor. Detalhamos agora algumas das mais importantes que estão implementadas em sistemas de *streaming* para transmissão na Internet:

III.2.1 Escalabilidades do tipo SNR, espacial e temporal

Uma das formas de adaptar um vídeo transmitido na Internet às condições específicas de transmissão ou das capacidades dos receptores é fazer diminuir o tamanho da imagem apresentada resultando na escalabilidade espacial, reduzir o número de imagens por segundo, o que leva à escalabilidade temporal, ou utilizar a escalabilidade SNR fazendo empobrecer propositadamente a qualidade da imagem.

Escalabilidade SNR (Signal-to noise ratio)

Nesta técnica, a codificação de vídeo produz subrepresentações que têm a mesma resolução espacial e a mesma taxa de imagens por segundo mas que utilizam um número

diferente de níveis de quantificação. A camada base é a que utiliza mais níveis mas pode ser conjugada com camadas de melhoria para a obtenção de um melhor resultado.

Escalabilidade Espacial

Neste caso, tanto a camada base como as camadas de melhoria têm iguais parâmetros no número de imagens por segundo e na qualidade de quantificação mas diferentes resoluções espaciais. As camadas de melhoria têm uma menor resolução espacial mas se forem adicionadas à camada base fazem aumentar o tamanho da imagem do vídeo.

Escalabilidade temporal

No último caso, as camadas de melhoria têm uma menor taxa de imagens por segundo do que a camada de base, enquanto que os níveis de quantificação e de resolução espacial são idênticos. As camadas de melhoria têm por função entregar ao decodificador as imagens não contidas na camada base para assim se obter uma melhor desempenho do sistema.

Em qualquer destes 3 sistemas, apenas se regista uma melhoria de qualidade no vídeo se pelo menos uma das camadas de melhoria for integralmente recebida. Como resultado, há um número muito limitado de níveis possíveis para a qualidade do *streaming* o que pode ser considerado manifestamente pouco na tentativa de rentabilizar as capacidades dos canais de transmissão de maior débito. Estes tipos de compressão, contudo, demonstram uma grande eficiência quando se trabalha sobre um canal estável e se afinam os parâmetros da compressão para essas características. Este tipo de compressão pode ser visto como o “Escalável 1” do gráfico III-2. No entanto, a Internet actual não pode oferecer um canal de comportamento estável e por isso estas técnicas são preteridas em função de outras a que podemos denominar progressivas e que serão apresentadas a seguir.

III.2.2 Fine Granularity Scalability (FGS)

Com o intuito de providenciar um melhor desempenho e flexibilidade no que diz respeito à resposta a diferentes tipos de ligação dos clientes, surgiu uma técnica de compressão escalável denominada *Fine Granularity Scalability* (FGS) [3]. Esta técnica surge associada à norma MPEG-4 e é uma das ferramentas que permite a adaptação da tecnologia aos diferentes ritmos de transmissão da Internet.

O FGS também utiliza uma camada base e uma ou mais camadas de melhoria mas a grande diferença desta técnica é que qualquer porção das camadas de melhoria pode ser utilizada para trazer melhorias à camada base. Ao contrário das técnicas vistas no ponto atrás, não é obrigatória que toda a camada seja recebida pelo decodificador de *streaming* para poder ser usada. Assim, um sistema que receba, por exemplo, 10% desta camada terá um pequeno melhoramento na qualidade do vídeo em relação ao sistema que só absorveu a camada base. No entanto, outro sistema pode receber 50% da camada e ter melhor desempenho e assim sucessivamente. O resultado é um método capaz de ter ajustes finos nos níveis de qualidade, ou seja, muito granular neste aspecto, daí advindo o seu nome. Se voltarmos ao gráfico III-2, estaremos perante a linha correspondente ao “Escalável-2” e que se aproxima ao ideal pretendido.

Modo de funcionamento do FGS

A construção da camada base em FGS é feita sem qualquer recurso a codificação escalável. A obtenção desta camada é feita pelo método da aplicação da Transformada Directa de Coseno (DCT) à sequência resultante da diferença entre as imagens consecutivas, tal como definido pela norma MPEG, e o objectivo é a construção de uma bitstream capaz de ser lida por um utilizador com uma ligação que seja considerada mínima segundo os objectivos do fornecedor de *streaming*. Está preparada para responder aos parâmetros de perda de pacotes e atrasos numa ligação de fraca qualidade.

A diferença na qualidade de vídeo entre este tipo de utilizador e aquele que possui os melhores recursos na ligação e nas capacidades do decodificador é proporcionada pelas camadas de melhoria. Estas camadas, que podem ser cortadas em qualquer quantidade de bits, acabam por ser as responsáveis pela escalabilidade desta técnica.

Uma camada de melhoria obtém-se por codificação da diferença entre a imagem inicial e a imagem reconstruída quando se usa codificação dos coeficientes DCT pelo método bit-plane [3]. Este método pode sofrer pequenas variantes na sua estrutura, nomeadamente na construção da camada de melhoria através da codificação por bitplane, da diferença entre os coeficientes DCT da imagem original e os coeficientes DCT da desquantificação da camada base [5]. Seja qual for o método utilizado, e uma vez que cada camada se apoia na camada base, as camadas de melhoria também servem para a correcção de erros da transmissão e, por outro lado, erros que ocorram durante a transmissão da camada de melhoria também não afectam a qualidade das imagens seguintes.

O número de camadas de melhoria em FGS não é fixo. É igual ao número de bitplanes necessários para a codificação na obtenção das mesmas. Cada uma tem mais detalhes para a reconstrução das imagens pelo decodificador.

Um esquema possível para o codificador FGS é o mostrado na seguinte imagem:

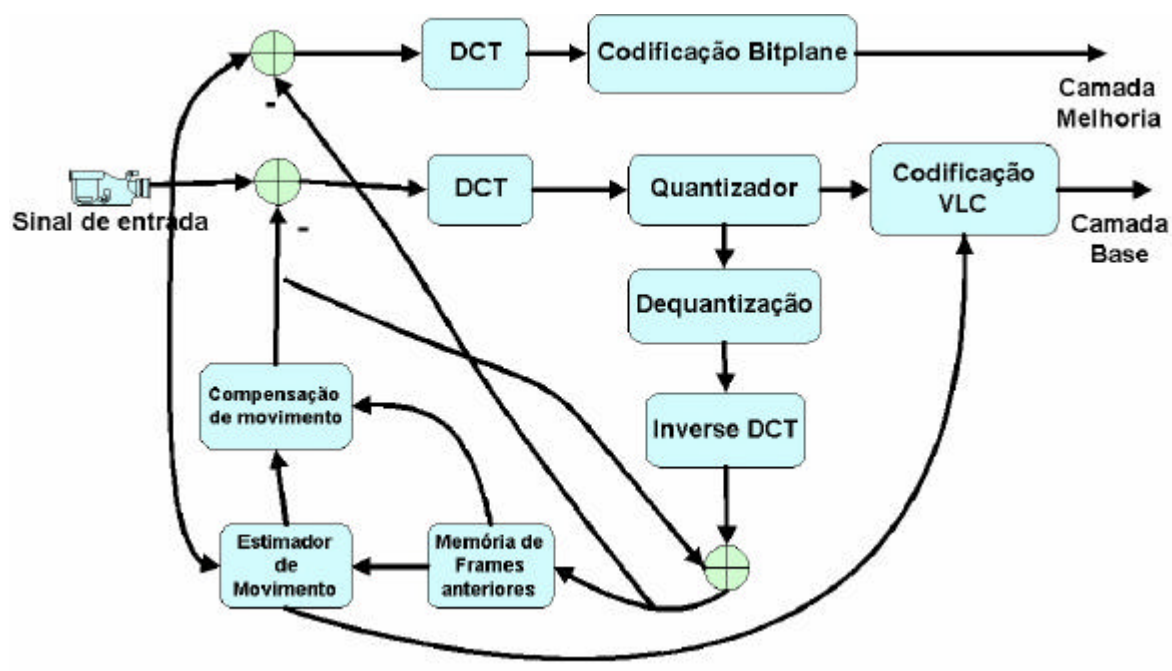


Fig III-3 – Codificador FGS

No lado do decodificador há ainda um outro patamar de escalabilidade pois é possível ao decodificador voltar a cortar a camada de melhoria se a sua capacidade de processamento for ultrapassada. Desta forma, para além da escalabilidade em relação às condições de transmissão obtemos também escalabilidade em relação às capacidades da máquina onde está instalado o leitor de *streaming*.

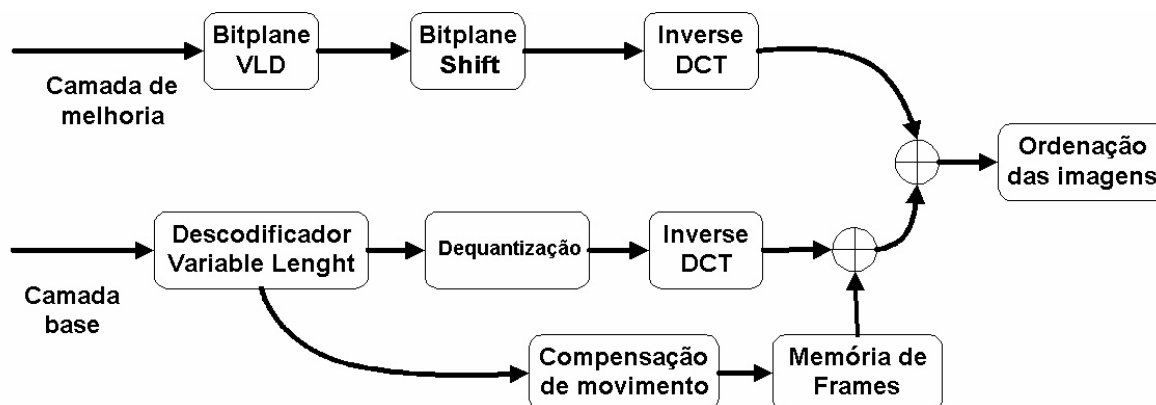


Fig. III-4 – Decodificador FGS

Para além das vantagens que advêm das características atrás exibidas, a própria simplicidade do codificador e do decodificador ajudam à implementação desta técnica.

As várias análises críticas à FGS [6 e 7] demonstraram que se atingem os objectivos de adaptação ao canal de transmissão e recuperação de erros, mas que se perde eficiência de compressão se comparada com técnicas do tipo SNR desenvolvidas para uma condição específica.

III.2.3 Progressive Fine Granularity Scalability (PFGS)

Na técnica FGS, as camadas de melhoria são sempre construídas a partir da camada base. A PFGS surgiu com base na ideia de construir camadas de melhoria baseadas também nas camadas de melhoria antecessoras e assim fazer com que cada uma traga mais precisão na reconstrução do bitstream inicial para benefício da qualidade final da sessão de vídeo na Internet e assim aumentar a eficiência de codificação [5].

Modo de funcionamento do PFGS

A camada base é obtida da mesma forma que na FGS. A diferença é que uma camada de melhoria se constrói com base numa outra e não somente pela camada base.

Como a camada base é codificada com um baixo bitrate, a determinação da compensação de movimento tendo-a só a ela por referência pode ser muito imprecisa. Assim, como a camada de melhoria possui maior qualidade que a camada base, a sua utilização na construção de uma nova resulta numa camada ainda mais qualitativa e mantém quer as características de granularidade como também a possibilidade de correcção de erros dada a correlação existente entre as várias camadas.

Em relação à FGS, um sistema PFGS necessita de mais buffers, pois para além da camada base, é necessário guardar informação das várias camadas de melhoria para se construírem

as seguintes o que faz aumentar a capacidade de processamento dos codificadores e aumentar o custo de construção dos mesmos.

Hoje em dia, os utilizadores de PFGS tentam atingir um compromisso entre o custo da operação e a eficiência de codificação. Por isso, em vez de se usarem como referência todas as camadas de melhoria, opta-se por utilizar aquelas onde a probabilidade de erros seja menor. As camadas de menor nível são tipicamente rejeitadas pois podem introduzir maiores erros quando há grandes movimentos entre *frames* consecutivas. As camadas mais altas também não servem pois exigem maiores taxas de transmissão que podem não ser atingíveis pela maior parte das ligações. Sobram as camadas intermédias cujos coeficientes DCT mostram ser altamente correlacionados em imagens adjacentes.

Tipicamente, os codificadores PGFS utilizam 2 buffers:

- ? Um para guardar a camada base
- ? Outro para guardar uma camada de melhoria reconstruída para servir de referência a camadas superiores

Testes feitos até agora com a técnica PFGS [5] revelam ganhos de eficiência de codificação em relação à FGS, tal como tinha sido previsto inicialmente. As características de escalabilidade granular e recuperação de erros não são afectadas. Em situações de comparação com técnicas do tipo SNR em canais estáveis, estas ainda levam vantagem.

Tem de ser levado em conta que a crescente complexidade dos decodificadores para esta técnica pode ser um óbice aos novos meios de ligação à Internet, nomeadamente via telemóvel ou PDA, pois são aparelhos que, pelo seu tamanho, não se coadunam com sistemas de maior complexidade e de necessidade de memória e recursos computacionais.

III.2.4 Técnicas de compressão híbridas Temporal/FGS

O método FGS proporciona adaptação de qualidade de imagem à capacidade instantânea do canal de transmissão mas não proporciona melhorias na taxa de imagens por segundo.

O objectivo das técnicas híbridas é conjugar a granularidade de qualidade de imagem do FGS com a capacidade de adaptação à taxa de imagens por segundo das técnicas do tipo escalável Temporal [4].

Para isso utiliza-se uma técnica que utiliza uma única camada de melhoria e adapta-se o sistema FGS para introduzir uma componente de escalabilidade SNR [4]. Esta camada de melhoria pode sofrer alterações no ritmo de transmissão em função da capacidade do canal para suportar blocos de informação que permitem aumentar o número de imagens entregues ao decodificador. A codificação passa a ser independente do processo de transporte para permitir adaptações de largura de banda ou preferências do cliente. No primeiro caso, adiciona-se um sistema pouco complexo que tenta determinar a taxa de transmissão instantânea para poder introduzir mais imagens por segundo na transmissão. No segundo caso, permite-se ao utilizador final a possibilidade de dialogar com o codificador para lhe pedir uma versão de maior qualidade do trecho de vídeo.

A pouca complexidade adicional deste sistema, permite a sua adaptação em aparelhos poucos potentes para melhorar o desempenho destes equipamentos numa sessão de vídeo em tempo real na Internet se comparados com a técnica FGS isolada.

III.3 Desenvolvimentos no controlo de Qualidade de Serviço

Os principais objectivos dos desenvolvimentos nesta área são o evitar o congestionamento e maximizar a qualidade do vídeo mesmo na presença de perdas de pacotes. O desenvolvimento das técnicas específicas para esta função foi feito de modo a não necessitar de qualquer intervenção ao nível da rede de transporte e por isso são implementadas nos emissores e receptores.

III.3.1 Controlo de congestionamento

Como já foi referido, para evitar congestionamentos na rede, o *streaming* utiliza técnicas de controlo de emissão, aumentando a taxa de transmissão em períodos leves e diminuindo em períodos sobrecarregados. Para atingir estes objectivos há dois principais mecanismos[2]:

- ? Controlo de taxa de transmissão
- ? Rate Shaping

Controlo de taxa de transmissão:

Baseia-se no controlo da informação enviada em função de uma estimativa da largura de banda disponível na rede. Há algumas variantes deste método, dependendo de quem é que determina a taxa de transmissão: o emissor, o receptor ou ambos. No primeiro caso, o objectivo é atingir um compromisso entre a quantidade de informação enviada e a perda de pacotes gerada; no segundo, o emissor descarta alguns canais provenientes de diferentes emissores (no caso de multicast) para poder tratar a tempo a informação a apresentar. Finalmente, se tivermos um modelo híbrido, há a conjugação dos dois factores.

Rate shaping:

Neste método, é no codec emissor que está a inteligência do processo pois cabe-lhe alterar a forma de codificação para a ajustar à taxa de transmissão permitida pela rede. Para o fazer, pode optar por eliminar frames ou ter filtros capazes de filtrar camadas de subrepresentação, de reduzir o número de coeficientes na codificação por DCT (o que provoca alterações na qualidade da imagem ao nível da cor que pode mesmo chegar ao ponto de termos uma transmissão a preto e branco).

III.3.2 Controlo de erros

Existem alguns métodos para reduzir os erros de uma transmissão de *streaming* [2]. Um deles consiste na introdução de pacotes redundantes na transmissão, dando assim ao receptor mais oportunidades de captar os pacotes pretendidos. Este método denomina-se *Forward Error Correction* (FEC). Outra forma é a de fazer a retransmissão dos pacotes perdidos, o que muitas vezes não é possível, pois o tempo de viagem deste pacote retransmitido é geralmente superior ao tempo que a apresentação do vídeo pode esperar. Contudo, quando o tempo de viagem é curto, este método torna-se viável e é denominado por *Delay-Constrained Retransmission*. Outra forma, denominada *Error-Resilient Encoding*, tenta tornar o formato do vídeo comprimido mais robusto à perda de pacotes. Pode-se fazer isto utilizando uma codificação com múltiplas descrições do objecto, cada uma capaz de reproduzir, com qualidade limitada, a fonte. Consoante as características do

canal, o receptor pode receber sem erros mais representações melhorando a qualidade da apresentação. Por último, um método denominado *Error Concealment* é implementado pelo receptor em reacção à percepção de que se perdeu um pacote substituindo-o por um baseado na informação dos últimos pacotes correctos (informação temporal) ou pela observação da imagem e dos pixels vizinhos (informação espacial).

III.4 Desenvolvimento de serviços de distribuição contínuos

Neste ponto vamos analisar o que pode ser feito ao nível da rede para melhorar o *streaming* de vídeo na Internet. Construídos em cima do protocolo IP, estes serviços de distribuição contínua são desenhados para proporcionar Qualidade de Serviço e tornar possível o *streaming* de vídeo numa rede do tipo Best-effort.

Temos três tipos de serviços [2]:

- ? Filtragem pela rede
- ? Multicast do nível de aplicação
- ? Replicação de conteúdos

Passamos a fazer uma descrição sumária de cada um.

III.4.1 Filtragem pela rede

Geralmente os ISP's colocam filtros na rede cuja função principal é adaptar a taxa de transmissão dos troços de rede que lhe são contíguos às características de congestionamento desses mesmos troços. Cada filtro pode actuar bidireccionalmente, ou seja, nos pacotes de *streaming* que recebeu e que terá de enviar ao cliente final ou ao próximo filtro, ou nos pedidos dos clientes ou filtros que terá de transmitir para montante. Este método usa também as medições de perdas de pacotes em comparação com um limiar para controlar a taxa de transmissão.

Assim, podemos atingir uma boa eficiência da utilização da largura de banda disponível para que a qualidade do vídeo apresentado seja a melhor possível dadas as circunstâncias.

III.4.2 Multicast do nível de aplicação

A tecnologia IP Multicast é uma tentativa de se implementar sobre a camada IP a entrega de pacotes para vários receptores que seja mais eficiente. Em vez de se criar uma transmissão dedicada desde o emissor até ao receptor, tenta-se que em cada troço da rede haja apenas uma versão de um evento que possa ser posteriormente adaptado nos troços finais a um utilizador em concreto. Por exemplo, nas duas figuras seguintes (Figs III-5 e III-6), temos um sistema de distribuição de vídeo constituído por um servidor, vários troços de rede (ligações mais grossas) e vários clientes.

Na primeira das figuras não existe Multicast. Para cada cliente, o servidor envia uma cópia (traços finos), o que faz com que em cada troço possam existir várias cópias do vídeo. Por

exemplo, entre o Router A e o Router C, há 4 cópias em simultâneo o que representa um grande desperdício de largura de banda.

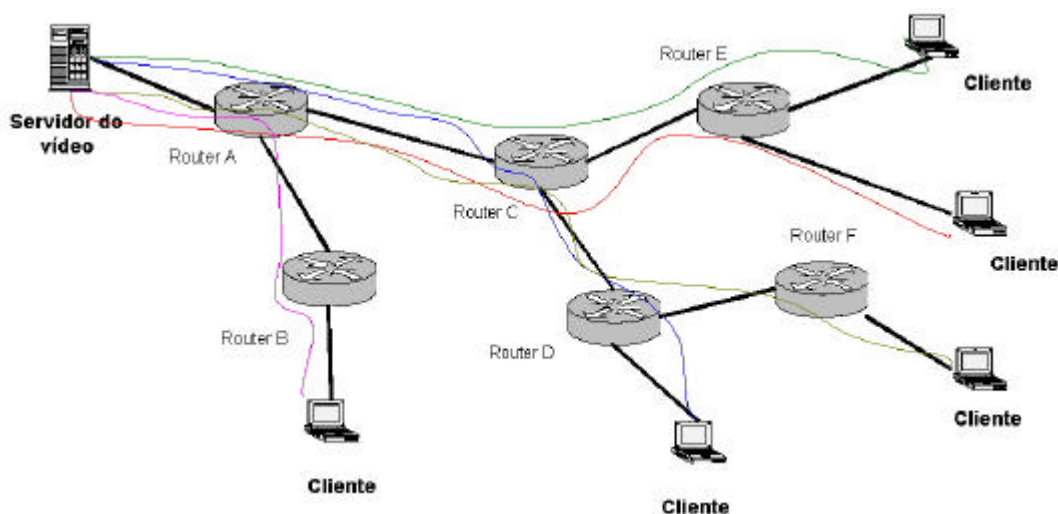


Fig III-5 – Distribuição de Vídeo sem Multicast pela Rede

Se em cada troço da rede apenas uma cópia do pacote original for transportado, estaremos a rentabilizar sobremaneira a nossa rede. Esta implementação é muito benéfica para os ISP's, para os fornecedores de conteúdos na Internet e, por consequência, para os utilizadores finais.

Para a construção desta rede, colocam-se os nós de distribuição de Multicast designados MediaBridge, ou seja, nós de rede com a inteligência suficiente para determinarem em conjunto qual a melhor rota ao longo da rede para a distribuição de conteúdos através de algoritmos de routing, para além da função de verificarem se um determinado conteúdo está a ser pedido repetidamente por um outro MediaBridge que já o tenha recebido para assim se evitar a repetição do envio. Na próxima figura, temos um exemplo em que mesmo no troço entre a máquina A e a C, apenas existe uma cópia do vídeo:

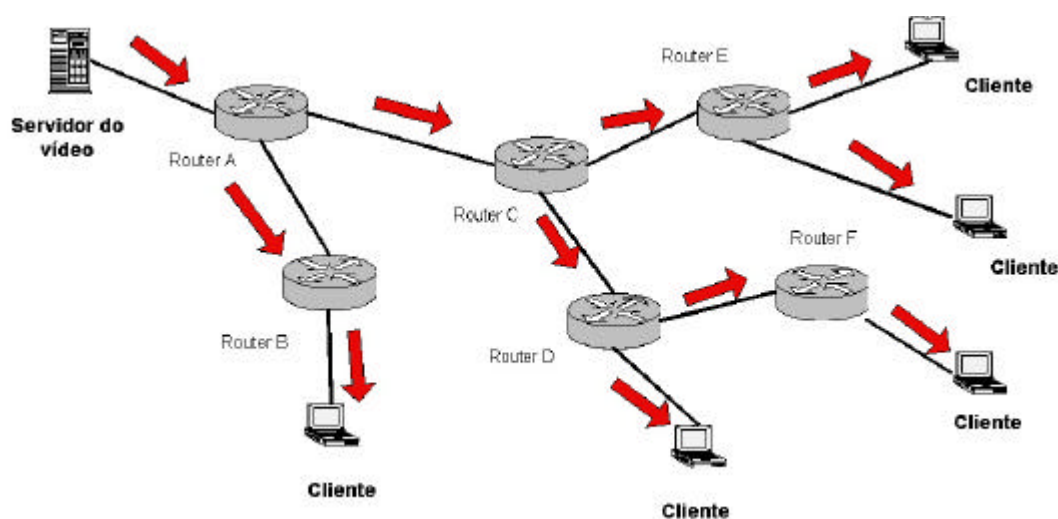


Fig III-6 – Distribuição de vídeo numa Rede com IP Multicast

Os benefícios são evidentes: rentabilização da largura de banda disponível, controlo de congestionamento e, como a transmissão de dados é independente do número de clientes finais, a rede sofre poucos incrementos de utilização.

Layered Video Multicast

Uma das técnicas desenvolvidas para permitir o Multicast de *streaming* de vídeo na Internet é o denominado Layered Video Multicast. A emissão de *streaming* em Multicast é feita, desde o emissor, em camadas que se diferenciam por permitirem várias larguras de banda da transmissão. Ao longo dos troços da rede, os Routers Multicast difundem uma cópia de cada camada.

Cada cliente adere a um grupo de Multicast solicitando ao ponto de ligação mais próximo de si a cópia da versão na camada que lhe convém. Esta técnica previu a escalabilidade da situação, solucionando os problemas de alteração de condições de largura de banda do utilizador ao permitir que um cliente possa, durante a transmissão, aderir a uma camada superior ou inferior de acordo com as alterações que lhe sejam subjacentes.

III.4.3 Replicação de conteúdos

A replicação de conteúdos é designada, nas suas duas vertentes, por *caching* ou *mirroring* e o seu objectivo é o de disponibilizar a informação o mais próximo possível dos seus destinatários.

No *caching*, cada vez que um cliente recebe um determinado conteúdo, um servidor guarda uma cópia dessa informação. Assim, quando um outro cliente fizer um pedido semelhante que tenha de passar por esta máquina, o pedido será respondido logo por ela e não pelo fornecedor original. Este processo é já vulgar mesmo ao nível das LANs de empresas ou instituições para rentabilizarem as suas limitadas ligações via os operadores. De outra forma, as necessidades de tráfego poderiam ser bastante superiores e originar maiores custos. Ao longo de toda a rede IP podem mesmo ser estabelecidas hierarquias de *caching*. Por outro lado, no *mirroring*, a iniciativa de se guardar uma determinada informação é tomada pelos gestores de rede. São implementadas máquinas específicas para este fim que terão periodicamente de actualizar junto da fonte a informação que guardam. Este processo é mais complexo do que o *caching* e é tecnicamente mais exigente, logo os custos associados são superiores.

Tanto um método como o outro tentam diminuir a sobrecarga da rede pela instalação de réplicas em locais específicos das suas redes. No caso de *streaming* vídeo, estes métodos podem ser usados idealmente para guardar toda a mensagem de vídeo mas está demonstrado que mesmo se forem replicadas algumas frames o efeito benéfico também é significativo.

III.4.4 Redes Distribuídas

A transmissão de *streaming* pela Internet, se feita apenas por um servidor, pode tornar-se redundante se dois clientes que partilhem troços de rede solicitarem a mesma informação. Ou seja, no mesmo link podem circular em tempos muito próximos, duas imagens do mesmo objecto o que é muito mau em termos de gestão de um dos recursos mais importantes da Internet que é a sua largura de banda. Como consequência, o troço que se liga ao servidor assiste a um enorme fluxo de pacotes redundantes o que pode provocar congestionamento.

Algumas soluções têm vindo a ser implementadas para melhorar o desempenho destes sistemas:

- ✍ Multicast Routing: Os Routers da rede guardam uma réplica da informação vinda do servidor que enviam a cada um dos clientes a jusante que a solicitem, funcionando assim como *cache*. Esta solução implica custos elevados em equipamento.
- ✍ Servidores intermédios: Esta solução é geralmente implementada pelos ISP's e consiste em colocar servidores a fazerem *mirroring* das informações mais solicitadas pelos seus clientes. Desta forma, limita-se bastante uma das componentes de maior custo para os ISP's que é a sua conectividade internacional ou com outros ISP's. Por outro lado, os clientes acedem à informação de uma forma mais rápida.
- ✍ Largura de Banda reservada: por vezes, para garantir um determinado grau de qualidade à transmissão de um evento, os ISPs reservam-lhe uma determinada largura de banda nos troços da rede entre os Routers. Esta opção tem custos limitados mas implica uma diminuição de capacidade para as restantes solicitações à rede.
- ✍ Transmissão de acesso múltiplo: uma outra solução é substituir o servidor único do *streaming* por um conjunto deles estrategicamente colocados na rede.

III. 5 Desenvolvimento de protocolos adaptados para Streaming na Internet

Perante o crescimento da tecnologia de *streaming* e da necessidade de contornar alguns problemas da Internet foram adaptados protocolos específicos para este cenário, escolhidos entre os vários standards usados. Os protocolos são específicos para cada camada do sistema, nomeadamente para os níveis de rede, transporte e sessão.

Podemos ver na seguinte figura (Fig III-7), a interacção dos diferentes protocolos envolvidos:

Sessão	Conference Control.		RTSP	SDP			
		RDP/RTCP		SDAP	FTP	HTTP	
Transporte	RSVP	UDP		TCP			
Rede	IP						

Fig. III-7 - protocolos numa sessão de streaming

III.5.1 Protocolo de rede

Neste caso, o protocolo IP é também o utilizado pelos sistemas de *streaming* sem qualquer modificação. Todas as operações de endereçamento são feitas de forma semelhante às restantes aplicações sobre a Internet. Convém realçar que o protocolo IP é independente do suporte físico e tanto se aplica a redes fixas como móveis ou sem fios. Por isso mesmo é

considerado como um ponto de convergência para as redes de transmissão de dados, voz ou multimédia.

III.5.2 Protocolos de transporte

Nesta camada, temos os protocolos UDP e TCP responsáveis pelo transporte da informação. Para além das características intrínsecas de cada um deles, convém referir que ambos implementam mecanismos de recuperação de erros mas o TCP utiliza a retransmissão como forma de o fazer. Desta forma, este protocolo provoca atrasos que muitas vezes não são aceitáveis pelas aplicações de *streaming* e por isso, tipicamente, o UDP é preferido. Contudo, como não é capaz de garantir a entrega da informação, os sistemas de *streaming* têm de implementar eles próprios mecanismos de recuperação de informação perdida. Para compensar a ausência de garantia de entrega de pacotes deste protocolo, é utilizado em cima de UDP o protocolo **Real Time Protocol (RTP)**, desenhado precisamente para suportar aplicações de tempo real na Internet no que diz respeito à detecção de perdas de pacotes. É o mais importante dos standards desenvolvidos para *streaming* pois todos os pacotes de *streaming*, seja qual for o formato, são encapsulados em pacotes RTP. Complementa o TCP pois o seu cabeçalho tem campos adicionais, entre os quais um marcador de tempo e um número de sequência para permitir a correcta reconstrução. Não actua por retransmissão de pacotes perdidos. É um protocolo de transferência de informação que pode ser complementado com o **Real Time Control Protocol (RTCP)**, que tem como vantagem fornecer informações de QoS aos utilizadores do RTP.

Algumas das funções garantidas pelo RTP são [2]:

- ? Marcação temporal da informação, o que pode ajudar à sincronização;
- ? Marcação de sequência, importante porque o UDP não garante uma entrega por ordem de envio;
- ? Marcação pelo tipo de payload, para identificação dos pacotes pertencentes aos diferentes trechos;
- ? Identificação da fonte que é útil nos casos em que o receptor recebe a informação de fontes distintas.

No que diz respeito ao RTCP, usado pelos utilizadores do RTP para trocar informações sobre a qualidade da transmissão, este protocolo implementa principalmente as funções de *feedback* de qualidade de serviço através do envio de determinados pacotes entre o servidor e o receptor onde são caracterizadas as condições da rede no que diz respeito a percentagem de pacotes RTP perdidos entre relatórios, percentagem de pacotes perdidos desde o início da sessão, diferença de atrasos nos pacotes ou o atraso desde a recepção do último relatório. Estas informações são usadas pela fonte, pelos operadores ou pelos receptores para a melhoria de desempenho da rede. Outras facilidades deste protocolo são: possibilidade da identificação dos participantes por um nome, número de telefone, e-mail ou outro; escalabilidade através de controlo de pacotes; sincronização inter-objects e outras informações sobre a sessão.

III.5.3 Protocolos de sessão

Ao nível de uma sessão de *Streaming* de vídeo em tempo real, o desenvolvimento de protocolos específicos tiveram como objectivo a identificação e o controlo das sessões.

O **Session Description Protocol (SDP)** é usado para a identificação da sessão através da atribuição de um nome específico, identificando o objectivo da mesma, os emissores e receptores e os endereços e portos usados por eles.

No caso de estarmos em Multicast, o protocolo **Session Announcement Protocol (SAP)** difunde pela rede informação acerca da sessão que serve de base aos clientes para se poderem ligar.

Para controlo de cada sessão, os protocolos usados pelo *streaming* de vídeo na Internet são o **Real Time Streaming Protocol (RTSP)** e o **SI Protocol (SIP)** [2].

O RTSP permite o controlo de sessões de Unicast ou Multicast em tempo real e possibilita a implementação das funções típicas de um VCR tais como pause, fast forward, etc., assim como o estabelecimento e controlo dos streams entre os servidores e os clientes, incluindo a função de informação de conteúdos adicionais disponíveis para transmissão. Por outro lado, cabe-lhe a escolha do melhor tipo de sessão para a entrega dos pacotes, passando de unicast a multicast ou até mesmo se isso for possível, implementando a sessão em cima de TCP.

Por seu lado, o SIP tem funções semelhantes mas também permite a mobilidade dos utilizadores fazendo de proxy para redireccionar os pacotes para a localização do utilizador.

III.6 Conclusão

Ao longo deste Capítulo, verificámos as várias vertentes que os investigadores na área de transmissão de vídeo têm feito para aperfeiçoar o *streaming* e contornar os obstáculos colocados pelas características da Internet. Ao longo de toda a cadeia do processo, desde o codec ao receptor passando pelos constituintes activos da rede, existem formas desenvolvidas especificamente para proporcionar aos utilizadores uma experiência mais agradável. Embora todos eles tenham importância e dêem o seu contributo, julgamos que há dois factores mais importantes: o aumento generalizado da largura de banda da Internet, quer ao nível dos troços de ligação, quer ao nível de *backbones* e ligações entre os ISP's, e os algoritmos de compressão, vistos em maior detalhe no Capítulo II.

Capítulo IV – O Futuro do Vídeo em tempo Real na Internet

Feito o retrato da actual situação das tecnologias de vídeo na Internet, cabe agora indicar as novas direcções que começam a tomar forma para o desenvolvimento da tecnologia.

Os passos a serem dados terão de acompanhar os novos desenvolvimentos técnicos e sociais que ocorrem de forma bastante célere no nosso tempo.

Do ponto de vista social, o desenvolvimento humano faz com que vários países em vias de desenvolvimento estejam ainda a despertar para as tecnologias aqui descritas. O exemplo mais significativo será porventura a China, onde reside um quinto da população mundial e que tem vindo a recuperar o atraso inicial. Podemos ainda verificar que nos locais onde a Internet já está consolidada, a própria apetência de estratos da população até aqui afastados destes mecanismos está a crescer e mais pessoas, algumas com poder de compra elevado, tendem a aderir às soluções implementadas via Internet. A chegada destes mercados emergentes poderá alterar radicalmente as necessidades a que têm de responder os fornecedores de conteúdos e obrigar a rede a fazer a respectiva distribuição sem constrangimentos.

No aspecto de novas tecnologias, estamos hoje em dia a assistir ao início da implementação comercial em larga escala da Internet como a conhecemos na rede móvel, baseada em *browsers* e acesso a páginas através do protocolo HTTP e outros. Trata-se aqui, também, de alargar o leque de potenciais consumidores de vídeo na Internet, sujeitos a condições de ligação bastante distintas das actuais formas de acesso à Internet. Também se verificam crescimentos de capacidade da rede mundial mas estes são praticamente absorvidos pelo contínuo crescimento do número de utilizadores e das necessidades dos já existentes. Por isso mesmo, é necessário que as actuais técnicas de *streaming* conheçam desenvolvimentos para a adaptação a este novo mundo.

Abordaremos neste Capítulo os desenvolvimentos técnicos que se estão a iniciar ou que ainda estão no âmbito da imaginação dos investigadores. São apresentadas 3 grandes áreas de desenvolvimento que têm implicações no vídeo na Internet:

- ? Desenvolvimento de novos serviços disponíveis na Internet
 - o Serviços Integrados - IntServ
 - o Serviços Diferenciados - DiffServ
- ? Novos desenvolvimentos da rede
 - o Crescimento da capacidade das máquinas que suportam leitores de *streaming*;
 - o Crescimento nas formas de acesso e capacidade dos circuitos, routers e servidores na Internet;
 - o Internet nas redes móveis;
- ? Aperfeiçoamento das técnicas adaptadas ao *streaming* de vídeo
 - o Análise ao desempenho do MPEG-4 parte 10 (H.264 ou Advanced Video Coding)
 - o Transcodificação

- MPEG-7
- MPEG 21
- Universal Multimedia Access

IV.1 - Novos serviços na Internet

Embora já existam técnicas para se reservar recursos para certas aplicações ou comunicações entre 2 utilizadores específicos numa rede IP do tipo *best-effort* como a Internet, a sua efectiva aplicação tem sido rejeitada pelos operadores pois isso significaria a existência de clientes privilegiados que partilhavam grandes recursos e outros, que tivessem de partilhar com os primeiros os mesmos troços de rede, que veriam o desempenho da sua ligação ser reduzida em função disso, aumentando a insatisfação e, provavelmente, dando origem a quebras no número de clientes e receitas.

Contudo, os protocolos adequados à reserva de recursos na rede poderão vir a ser uma das formas de implementação de vídeo em tempo real na Internet. O seu objectivo é proporcionar Qualidade de Serviço às aplicações em tempo real, diferenciando-as do restante tráfego circulante na rede e para o qual o *best-effort* é suficiente. Já há duas formas possíveis de implementar QoS numa rede IP, designadas por Arquitectura de Serviços Integrados e Arquitectura de Serviços Diferenciados [1].

Serviços Integrados - IntServ

O protocolo RSVP permite o estabelecimento de canais reservados entre dois pontos em redes IP. Antes da transmissão da informação, a aplicação de vídeo em tempo real tem de reservar recursos da rede e estabelecer um caminho fixo que não será alterado durante a transmissão. O emissor solicita ao Router mais próximo, através de uma mensagem PATH, a reserva de recursos para um tipo de tráfego específico. Esta mensagem de PATH será retransmitida de comutador em comutador e cada um devolverá ao antecessor uma mensagem RESV que confirma a aceitação do pedido e solicita a reserva de recursos. Consequentemente, é reservada largura de banda no troço e espaço nos *buffers* dos Routers. No entanto, as condições de um determinado troço podem inviabilizar a reserva de recursos e, nesse caso, é enviada uma mensagem de negação do pedido até ao emissor que cancelará a transmissão.

Existe portanto um mecanismo de admissão ao sistema de forma a impedir que haja reservas de recursos superiores à capacidade da rede.

No entanto, este tipo de abordagem requer da parte dos elementos activos de rede maior complexidade e mais funções pois, por exemplo, para além da implementação do protocolo RSVP, é necessário ordenar e priorizar os pacotes de dados consoante a qualidade de serviço exigida pela aplicação de que fazem parte. À escala da Internet, esta situação pode gerar uma quantidade de informação enorme a ser processada momentaneamente pelos routers e também necessidade de armazenamento de informação durante mais tempo o que leva a que não seja aconselhável a sua utilização no backbone Internet mas apenas nos troços finais de ligação ao utilizador.

Serviços diferenciados - Diffserv

Ao contrário dos sistemas de Serviços Integrados, em que há a obrigatoriedade de reserva de recursos e o estabelecimento prévio de um caminho ponto-a-ponto, os Serviços

Diferenciados actuam ao nível do escalonamento de prioridades de transmissão em cada comutador.

Foi estabelecido pelo *The Internet Engineering Task Force* (IETF) que se pode utilizar um dos campos do cabeçalho IP, neste caso o Type of Service, para indicar que se está a utilizar uma arquitectura de serviços diferenciados e indicar a prioridade solicitada. Cabe a cada elemento de rede ordenar os pacotes à saída em função do valor apresentado neste campo. Assim se eliminam grandes quantidades de informação necessárias à reserva de caminhos e que contribuíam para sobrecarregar de tráfego os troços de rede.

Contudo, para este esquema funcionar, o *Internet Service Provider* de cada utilizador terá de proporcionar este tipo de serviço, expresso geralmente num *Service Level Agreement* (SLA) [1] que define para cada utilizador ou aplicação a prioridade a estabelecer na rede. Obviamente que isto tem implicações ao nível comercial, pois será de esperar que um utilizador que beneficia desta capacidade terá de pagar mais ao seu ISP.

Há um número limitado de valores que podem ser utilizados no campo acima referido, o que significa que este tipo de serviço é mais escalável que o IntServ. Mas, por outro lado, apenas garantem um melhor tratamento a uma determinada aplicação e não um nível de desempenho estável. Se a rede estiver congestionada, o comportamento de uma aplicação com DiffServ apenas será relativamente melhor que as outras aplicações sem atribuições concretas de tempos de atraso ou perdas de pacotes.

IV.2 - Desenvolvimentos nos elementos de rede e formas de acesso

A par com o aperfeiçoamento de técnicas de compressão e transmissão de eventos em tempo real, a evolução contínua na capacidade dos suportes físicos é um facto que contribui para a difusão e facilidade de acesso a estas tecnologias.

Começamos por indicar o aumento de capacidade das máquinas que suportam leitores de *streaming* de vídeo. Mesmo estabelecendo como referência o momento em que a norma MPEG-4 ficou completa, os PC's, mesmo os portáteis, já registaram aumentos de capacidade de:

- ? Velocidade de processamento: hoje em dia é vulgar trabalharmos com processadores que já ultrapassaram a barreira dos GHz; O futuro Pentium 5 promete chegar ao 5 GHz;
- ? Memória RAM: outro dos elementos fundamentais para o tratamento das grandes quantidades de informação geradas no *streaming* de vídeo. Aqui, memórias de 1 GB já começam a ser padrão de mercado;
- ? Capacidade de armazenamento: um disco duro vulgar já tem 100 GB, espaço considerável para os amantes da transmissão de vídeo armazenarem cópias das emissões;

Olhando para o mercado emergente dos pequenos PDA's, as suas capacidades já começam a envergonhar os PC's de meados dos 90. Já trabalham com processadores a 400Mhz, memórias RAM 64, bem como ecrãs de alta resolução. Cada vez mais, conjugados com a capacidade de ligação à Internet, estes aparelhos representam uma forma de acesso à transmissão de vídeo em tempo real. Seguindo a lei de Moore, podemos esperar um rápido desenvolvimento destas capacidades. Quantos de nós não gostaríamos de poder assistir a

notícias em directo ou a eventos desportivos bastando somente um destes pequenos aparelhos. Este será de certeza um dos motores do desenvolvimento das técnicas de *streaming* num futuro próximo.

A evolução que mais rapidamente trará melhorias à transmissão de vídeo em tempo real na Internet é a da capacidade da rede. Esta evolução tem sido rápida nos últimos anos e tudo se conjuga para que poderá ainda ter grandes margens de progresso. Embora em Portugal ainda estejamos a assistir à consolidação da Banda Larga com o acesso por cabo e o ADSL, por outras paragens já se começam a aplicar outras tecnologias variantes do ADSL, genericamente designadas xDSL. Estas podem chegar a 5 vezes o desempenho máximo actualmente fornecido o que representa, por si só, a melhoria de uma das principais limitações de implementação de *streaming* tal como vimos nos Capítulos I e III. Ainda mais promissor, a recente proliferação das primeiras redes Wireless, utilizando a norma 802.11b, com funcionamento a 11Mbps no troço de utilizador representa um desafio aos operadores de Internet para o fornecimento de acesso a velocidades superiores. Já se encontra definida e bastante difundido o 802.11g, que possibilita uns fabulosos 54 MBps no troço final. Estarão os ISP's dispostos a transpor estas evoluções de capacidades para os seus backbones de forma a alimentar este desenvolvimento? Isto requer que os ISP's reforcem os seus backbones, generalizando a utilização das últimas tecnologias de transmissão em fibra óptica e que progressivamente aumentem as capacidades dos Routers e servidores de acesso a aplicações de *video on-demand* por exemplo.

Há ainda a considerar a esperada e sucessivamente adiada revolução nas redes móveis. Hoje em dia, já se conseguem fazer os primeiros passos para a transmissão de mensagens multimédia contendo pequenos trechos de vídeo. Ora este é o embrião de algo que se pretende no futuro a médio prazo com a definitiva implantação do UMTS ou da já falada 4ª geração.

Se tivermos a capacidade de nos ligarmos à Internet de forma estável com, pelo menos, uma velocidade idêntica à da rede fixa, e se conjugarmos este acesso com um aparelho capaz de implementar um leitor de *streaming* tal como está iminente nos PDA's, teremos mais uma frente de desenvolvimento alimentada por um público maioritariamente jovem que é apte a esta tecnologia. Basta dizer que em Portugal já se assistiu à transmissão de um evento desportivo na rede móvel para vermos o quanto já se progrediu mesmo sem o UMTS.

IV.3 - Novas técnicas para streaming de vídeo na Internet

Ao nível de técnicas de compressão vários desenvolvimentos estão em curso para melhorar o rendimento dos actuais sistemas de *streaming* na Internet. O AVC / H.264 que foi apresentado no capítulo II, permite já hoje importantes avanços e mostra caminhos para o futuro próximo desta tecnologia.

Eficiência e Utilização do AVC na Internet

A maior eficiência do método de codificação do AVC é um importante argumento para quem está limitado pelas taxas de transmissão como é o caso da maioria dos utilizadores de vídeo em tempo real na Internet.

Por exemplo, a transformada com coeficiente inteiros, introduz um ganho de eficiência de 5.38 dB quando comparada com a DCT. Resultados de testes feitos [10 e 14] revelaram que o AVC representa uma melhoria de desempenho de 39% em relação ao MPEG-4 Visual, de 24% em relação ao H263++, de 47 a 49% em relação ao H.263 base e de 50 a 64% em relação ao MPEG-2.

No entanto, esta evolução pode não chegar a todos os utilizadores, não devido às velocidades de transmissão, mas sim porque exige maior complexidade aos decodificadores [10 e 11]. Por exemplo, um decodificador que implemente todas as funcionalidades do AVC para obter os ganhos de eficiência na compressão acima indicados, é 4 vezes mais complexo que o decodificador para o MPEG-2 equivalente. Ora isto obriga a máquinas mais potentes, com maiores capacidades de processamento e de armazenamento e consequentemente maiores recursos energéticos. Se bem que os PC's estejam a evoluir nesse sentido, o AVC poderá não ser ainda a ferramenta indicada para utilizadores que estão ligados à Internet com PDA's ou com telemóveis mais avançados onde as limitações de processadores e baterias se tornam importantes.

IV.4 – Acesso Multimedia Universal, Escalabilidade e Transcodificação, MPEG-7 e MPEG-21

Depois de os métodos para transmissão de vídeo terem evoluído de forma bastante rápida nos últimos anos, a comunidade de investigadores e utilizadores pretende dar mais um passo importante e integrar todas as tecnologias existentes para colocá-las à disposição dos utilizadores de uma forma transparente.

O conceito de Acesso Multimedia Universal (**Universal Multimedia Access** ou UMA) define um conjunto de tecnologias que permitam, para qualquer utilizador situado em qualquer parte do mundo, em qualquer momento, que utilize qualquer tipo de leitor e qualquer forma de ligação aceder a qualquer tipo de conteúdo multimédia de forma transparente e adaptada às suas condições [9].

Este objectivo é certamente muito ambicioso tendo em conta a realidade actual que se caracteriza por múltiplas formas de acesso (Móveis, fixas, banda larga, banda estreita,...), diferentes normas para a representação dos conteúdos, emergência do consumo interactivo, etc. Por isso, é imperativo o tratamento adequado da informação, catalogando-a para uma perfeita identificação do seu conteúdo e estruturando-a para um mais fácil manuseamento. Assim, será imprescindível a existência de formas de troca de informação entre clientes finais e a rede para que seja possível a transmissão individualizada e adaptada às condições próprias. Com este conhecimento, poderão existir mecanismos de adaptação localizados nos servidores, nos equipamentos activos de rede ou em servidores intermédios como exemplifica a imagem IV-1:

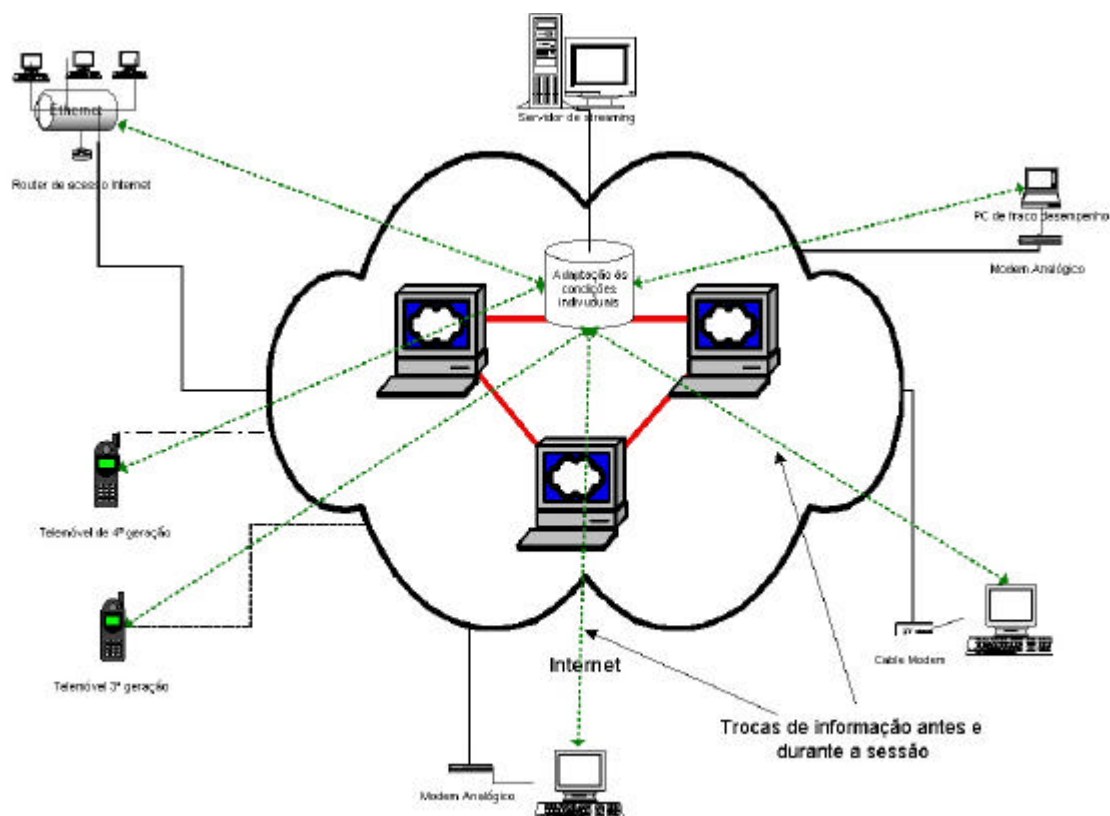


Fig. IV-1 – Esquema geral do conceito Universal Multimedia Access

As técnicas começam agora a surgir e podem ser divididas em 3 grandes grupos:

1. Ferramentas de adaptação às condições do indivíduo, principalmente a escalabilidade e a transcodificação;
2. Ferramentas para a completa descrição de um objecto, entre elas o MPEG-7;
3. Ferramentas para a completa descrição das condições específicas de um indivíduo para lhe serem atribuídas a versão do objecto que melhor se adapta a elas, entre as quais o MPEG-21;

Escalabilidade e Transcodificação

A escalabilidade da representação de um objecto é fundamental para a adaptação à taxa de transmissão de uma ligação na Internet. Nos Capítulos anteriores deste trabalho já foi plenamente identificada como uma necessidade e foram descritos para cada técnica de codificação as suas implementações para que seja uma realidade.

De facto, o MPEG-4 foi particularmente importante para a obtenção de escalabilidade, permitindo o acesso a conteúdos de vídeo em tempo real mesmo para ligações de baixo débito ou a aparelhos limitados na resolução espacial e temporal.

Outra das evoluções possíveis para a boa transmissão de vídeo em tempo real na Internet é a que estuda a possibilidade de a rede se tornar inteligente ao ponto de por ela própria ser capaz de mudar entre diferentes métodos de codificação do objecto, escolhendo para cada

troço ou para cada cliente aquele que melhor se adapta às suas condições, o que se designa por Transcodificação [7].

Se bem que a norma MPEG já inclui grandes possibilidades neste aspecto, muito por causa das características de escalabilidade do código utilizado, pretende-se dar um passo mais à frente e incluir elementos activos que saibam interpretar vários códigos diferentes, que conheçam o ambiente de transmissão no próximo encaminhamento e que traduzam para diferentes formas de codificação o bitstream recebido.

Deve ser evitada a forma primitiva de transcodificação que consiste em descodificar o sinal recebido por completo até obter a imagem inicial e depois voltar a codificá-lo com o método escolhido para a retransmissão. Este método, para além de consumir recursos importantes ao nível do processamento, ser dispendioso por obrigar a implementar múltiplos sistemas numa mesma entidade, exige sistemas complexos, podendo ser impraticável em sistemas de vídeo em tempo real por introduzir atrasos.

A construção de um sistema de transcodificação terá sempre o objectivo de obter eficiência na transformação minimizando as perdas de qualidade.

Entre as possibilidades destas técnicas temos:

- ? Adaptação a diferentes resoluções espaciais, para transmissão broadband para clientes que usam diferentes aparelhos leitores;
- ? Adaptação a diferentes resoluções temporais, para transmissão para leitores com diferentes capacidades de processamento;
- ? Mudança de débitos binários variáveis para débitos binários constantes na mudança de redes com diferentes comportamentos;
- ? Adaptação a ambientes com diferentes taxas de erros de transmissão, por exemplo na transmissão via Internet para aparelhos móveis
- ? Variações no formato de codificação: por exemplo de MPEG-2 para MPEG-4;
- ? Alterações de parâmetros de codificação como por exemplo o número de graus de quantificação;

Para ilustrar este conceito, a Figura IV-2 exemplifica 2 tipos de transcodificação. No primeiro caso, temos uma adaptação a novas condições de transmissão. No segundo, uma alteração de codificação para que um receptor com um tipo de leitor não adaptado ao sinal original possa aceder ao conteúdo multimedia.

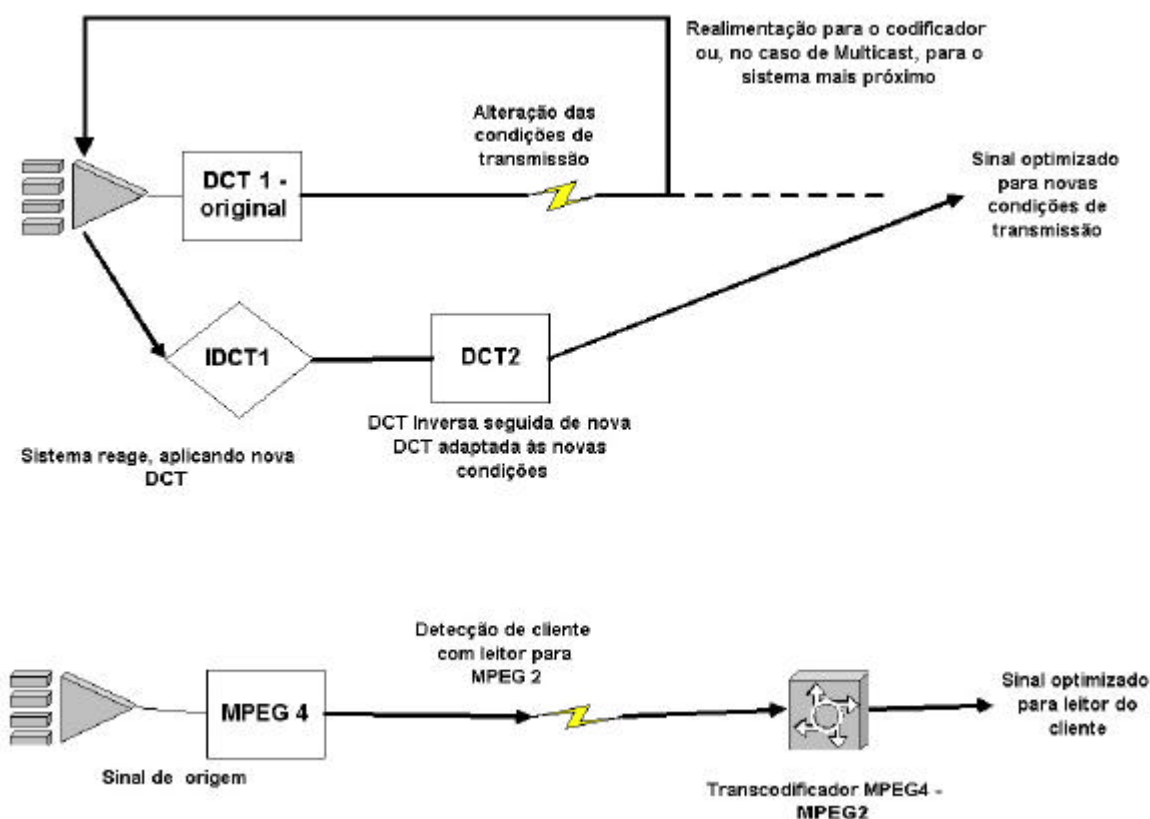


Fig. IV-2 – Transcodificação para adaptação a diferentes condições de transmissão ou para necessidade de alteração do tipo de codificação.

MPEG-7

A facilidade dos meios de produção, de armazenagem e de disponibilização de conteúdos de multimedia permitiu nos últimos anos um crescimento exponencial da quantidade dos mesmos na Internet. Por isso, é necessário que hajam formas de se lhes aceder de forma inequívoca. Ou seja, no meio da multidão de vídeos disponíveis na Internet, temos de ter a capacidade de encontrar especificamente o que procuramos e evitar a busca extensiva entre milhares de vídeos com conteúdos aproximados.

A resolução deste problema foi também pensada pelo Moving Pictures Experts Group. Foi constituído um grupo de trabalho com o objectivo de criar um método normalizado para a descrição de conteúdos para facilitar a busca de informação específica.

Assim nasceu o MPEG-7, denominado oficialmente por “*Multimedia Content Description Interface*”. Com a implementação desta norma, somos capazes de rápida e eficientemente descobrir, por exemplo, a transmissão de um noticiário em português sobre a descoberta de um novo método para tratar cancro da pele e cujo apresentador vista uma gravata vermelha, sem antes termos de pesquisar entre todos os canais de notícias em português na Internet, ou entre todas as notícias sobre cancro desse dia ou sem termos sido levados a um documentário sobre gravatas vermelhas.

Não se trata de uma nova norma de codificação mas sim da forma de associarmos a cada representação de um evento, feito em qualquer tipo de codificação, um conjunto de descritores, textuais ou não, obedecendo a parâmetros bem definidos que facilitem a busca

a um leitor capaz de interpretar essas indicações. Por exemplo, enquanto que o MPEG-4 define um standard para a representação das imagens do jogo acima referido, o MPEG-7 define como deve ser feita a informação sobre esse mesmo conteúdo ou seja a sua Metadata [6].

MPEG-7: Como funciona

A norma definiu um conjunto de ferramentas para a troca de informação sobre os conteúdos e mesmo de pormenores presentes no próprio vídeo, como por exemplo a marca da bola usada no jogo acima referido.

Tudo se baseia em **Descritores** e **Esquemas de Descritores**[6]. Um Descritor não é mais do que um texto, elaborado em linguagem XML, que obedece a uma estrutura previamente definida. Este texto inclui as informações que podem servir de base a respostas de pesquisas dos utilizadores. Um Esquema de Descritores não é mais do que a organização em blocos dos muitos Descritores para a construção de perfis de pesquisa ou perfis do utilizador.

A norma define ainda um método de compressão, de transporte e de armazenamento do próprio bitstream associado aos Descritores.

Paralelamente, o MPEG-7 permite que os programas utilizados para pesquisar um objecto, não definidos pela norma que apenas determina a estrutura da informação que o descreve, possam elaborar um sistema de perfis associado a cada utilizador ou mesmo a uma comunidade de utilizadores, quer por introdução do próprio das suas preferências, quer pela construção de um histórico sobre as tendências dos objectos requisitados. A norma define explicitamente um Esquema de Descritores denominado *UsageHistory* para albergar o histórico das acções associadas ao utilizador e um outro, o *UserPreferences*, onde se alojam as informações das preferências, indicadas pelo utilizador ou induzidas pelas suas actividades, como por exemplo uma lista de realizadores de cinema favoritos ou o clube de futebol que ele mais vezes pede para ver.

Para os utilizadores da Internet tem especial interesse a informação que é guardada e que diz respeito à capacidade da ligação à Internet e da máquina que este utiliza. Desta forma, o formato e a qualidade do vídeo que lhe será entregue é automaticamente ajustada sem a necessidade de escolha explícita do indivíduo.

Os conteúdos audiovisuais são tratados pelo MPEG-7 de forma a criar Sumários que os permitem descrever univocamente [6]. Este conceito pode mesmo ir ao ponto de segmentar um vídeo e associar-lhe sumários específicos que indicam as propriedades desses próprios segmentos. Dessa forma, permite-se uma busca mais fina e rápida a um determinado ponto de um objecto. Existem mesmo métodos automáticos para criar segmentos de vídeo que funcionam na base da pesquisa de descontinuidades na imagem que permitam identificar uma mudança de capítulo dentro de um filme, por exemplo. A cada um dos segmentos é associado um sumário que contém informações sobre o filme de onde foram retirados mas também de características desse mesmo segmento. Deve notar-se que estes sumários podem mesmo ser conjuntos de imagens retiradas do segmento e que dão ao utilizador uma visão rápida e concisa do trecho. Ora isto é muito importante para sistemas de difusão de vídeo-on-demand, dado que quando um utilizador pretende escolher um de vários filmes, ele selecciona-os visualizando precisamente o Sumário em MPEG-7. A figura IV-3 tenta mostrar, de forma genérica, como o processo se pode implementar.

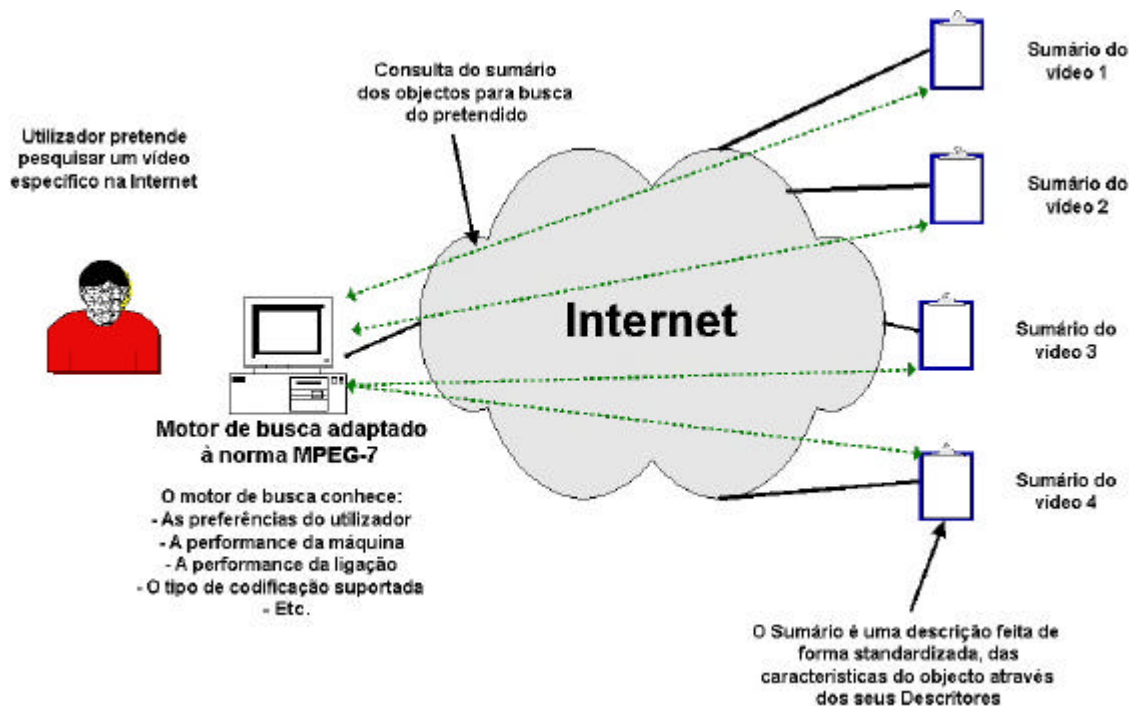


Fig. IV-3 – Esquema geral para um sistema com MPEG-7

Embora não seja um método que permita facilitar a transmissão de vídeo em tempo real na Internet, o MPEG-7 será sobretudo uma ferramenta inseparável dos utilizadores desta tecnologia, ao facilitar-lhes a pesquisa do que pretendam, poupando-lhes tempo e recursos. Poderá ainda tornar mais rápida e eficiente a adaptação da qualidade do vídeo, ou do formato de codificação do mesmo às condições específicas de cada um. Está definido um Esquema de Descritores chamado *Media Transcodification Hints* cuja função é a de conhecer as características do meio de transmissão e a de dar informações aos servidores ou a elementos activos como os gateways para a adaptação da imagem final a essas características ou mesmo às preferências do receptor, nomeadamente indicando uma adequada transcodificação.

Este é um contributo importante para a evolução do vídeo na Internet.

MPEG-21

Ao longo deste trabalho foram identificados múltiplos desenvolvimentos para a tecnologia de transmissão de multimédia na Internet. Depois de termos definido tantos métodos para a codificação, para o transporte, para a identificação dos elementos multimédia e de termos referido que existem numerosas formas de ligação à rede, é necessário criar uma forma de integrar todos estes elementos de forma transparente ao utilizador para que verdadeiramente se possa falar de Universal Multimedia Access.

Mais uma vez o MPEG criou um projecto para a conjugação de esforços nesse sentido que foi denominado MPEG-21 “Multimedia Framework”, onde o 21 provém do objectivo de associar esta ferramenta ao novo século [8].

Pretende-se criar um modelo de interacção dos variados métodos existentes, dos seus actores, produtores ou consumidores, saber como se conjugam entre si, discutir a necessidade de criar outras normas para preencher as lacunas e assim fazer surgir um

modelo aberto para a modelização da entrega de forma transparente de conteúdos multimédia e assim promover o seu uso.

A visão do MPEG-21 é a definição de um sistema para a utilização transparente, eficiente e fácil dos recursos multimédia independentemente das condições da rede de transmissão ou do tipo de aparelho receptor [8]. Tal como foi definido no UMA, pretende-se atingir a universalidade em termos de espaço geográfico, em qualquer momento e em quaisquer condições de suporte tecnológico.

O projecto MPEG-21 está dividido nas seguintes partes:

- ? Parte 1: Visão, tecnologia e Estratégia: Identificação dos objectivos e caminhos a seguir pela norma
- ? Parte 2: Digital Item Declaration: Definição da estrutura de um Elemento Digital
- ? Parte 3: Digital Item Identification: Ferramentas para a identificação de um elemento Digital
- ? Parte 4: Protecção e Gestão da Propriedade Intelectual
- ? Parte 5: Linguagem dos Direitos de Expressão
- ? Parte 6: Dicionário dos Direitos da Informação
- ? Parte 7: Adaptação do Elemento Digital
- ? Parte 8: Software de Referência
- ? Parte 9: Formato dos ficheiros

Existem já mais algumas partes em fase de estudo que pretendem complementar o trabalho das partes anteriores.

A pedra basilar do MPEG-21 é o **Elemento Digital (Digital Item)** que é definido como um objecto em formato digital representado segundo uma norma, que possui uma forma estruturada para acesso às suas parcelas, perfeitamente identificado (ver ponto anterior sobre MPEG-7). Cada um destes elementos não é mais do que o objecto multimédia que se pretende transaccionar, por exemplo uma transmissão de vídeo em tempo real, ao qual se adicionaram elementos para o adaptar à estrutura do MPEG-21, nomeadamente através de um descritor que o identifica como um todo e às suas partes constituintes. Estes elementos são tratados, trocados e acedidos por **Users**, ou seja, os agentes interagem com o ambiente MPEG-21. Podem ser os consumidores, produtores, comunidades ou operadores.

Cada elemento será assim identificado por um **Digital Item Declaration (DID)**, ou seja, um documento baseado em XML que o descreve univocamente tendo por base o MPEG-7, dando indicações sobre a sua estrutura, o seu conteúdo e indicações sobre o tratamento que pode sofrer. No caso de *streaming* de vídeo, cabe ao DID fornecer as informações suficientes para permitir a configuração e adaptação do vídeo entregue ao cliente. Associado ao Digital Item, também estará o **Digital Item Identification**, que é a identificação do Digital Item baseada num sistema que garanta a unicidade da identificação de forma permanente genericamente sob a forma de URI (Uniform Resource Identifiers).

Depois de um elemento estar perfeitamente definido, terá ainda associado um **Digital Item Adaption**. Este componente permitirá que o Digital Item seja manipulado para se adaptar às condições de transmissão de rede, do aparelho terminal ou mesmo das preferências do utilizador. Este manuseamento poderá mesmo implicar a alteração do formato de compressão do ficheiro original, se, por exemplo, o receptor tiver indicado essa preferência.

Digital Item Declaration	Digital Item Identification and Description	Adaptação de Conteúdos	Propriedade Intelectual e direitos de autor	Informação sobre a rede e terminais	Representação do conteúdo
Declaração do objecto na norma MPEG-21	Descrição detalhada do objecto e seu conteúdo, baseado no MPEG-7	Ferramenta que permitirá a adaptação do conteúdo às condições de transmissão ou do utilizador	Descrição das condições associadas à propriedade intelectual do objecto e às permissões de alteração ou de utilização	Parte responsável por identificar e definir as condições de transmissão e os requisitos dos terminais receptores	Forma em que está representado o conteúdo como por exemplo a sua codificação

Fig. IV-4 – Sistematização dos componentes da norma MPEG-21

É também dada grande importância à temática da protecção dos direitos de autor. É objectivo da norma que haja uma eficaz protecção desses direitos ainda mais elaborada do que no MPEG-4. Prevê ferramentas adequadas para a autenticação e troca de informações sobre propriedade intelectual.

Em resumo, no futuro, os utilizadores de recursos multimédia como, por exemplo, um vídeo em tempo real na Internet, poderão aceder aos eventos na forma de Elementos Digitais, que serão identificados univocamente e rapidamente, sem terem preocupações acerca do tipo de ligação ou tipo de aparelho que possuem, pois sabem que subjacente ao MPEG-21 está todo um sistema que automaticamente adapta a transmissão do vídeo às suas condições particulares e que poderá mesmo fazer a transcodificação do elemento multimédia para um formato que o seu leitor possa interpretar.

O MPEG-21 está ainda na fase de estudos e desenvolvimentos prevendo-se que esteja em condições de publicação até ao fim de 2004.

Capítulo V – Experiências Práticas de transmissão de vídeo na Internet

Com o objectivo de verificar a funcionalidade de alguns dos assuntos tratados neste trabalho planeámos um conjunto de experiências muito simples, ao alcance de qualquer utilizador médio de ferramentas de transmissão de vídeo.

As nossas regras impunham utilizar ferramentas disponíveis na própria Internet e que não acarretassem custos para o utilizador ou que esse custo fosse diminuto.

V.1 – Hardware e Ferramentas de Software

Na figura seguinte apresentamos o esquema montado para o nosso conjunto de experiências:

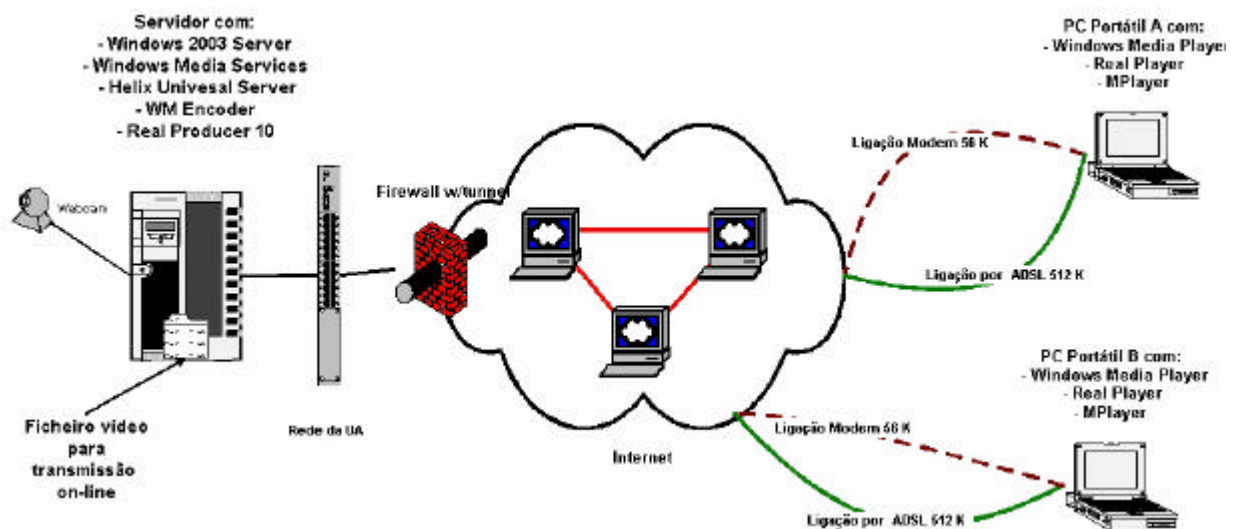


Fig. V-1 – Esquema da experiência prática

O objectivo deste sistema era obter resultados sobre a fiabilidade e capacidade de transmissão na Internet de um vídeo codificado a diferentes velocidades, fazendo variar também o programa codificador, a ligação do lado do receptor, as capacidades do PC onde residiam leitores de vídeo.

Para a codificação do vídeo a transmitir utilizámos duas ferramentas disponíveis gratuitamente na Internet:

1. Windows Media Encoder 9 [1]
2. Real Producer 10 Basic [2]

Cada um destes codificadores permite fazer o tratamento das imagens recolhidas por uma câmara de vídeo ou tratar um vídeo já existente e produzir diferentes níveis de codificação através da variação dos parâmetros da taxa de imagens por segundo, tamanho das imagens e taxa de transmissão da sessão de *streaming*.

Estes dois programas foram instalados numa máquina residente na rede do IEETA que por sua vez está ligada à rede da Universidade de Aveiro. O sistema operativo utilizado foi o Windows 2003 Server.

Para cada um dos codificadores, decidimos utilizar os servidores da mesma gama de produtos. Assim, para o Windows Media Encoder utilizámos os Windows Media Services [3], enquanto que para o Real Producer escolhemos o Helix Universal Server [4].

Um dos parâmetros que mais influência poderia ter no desempenho deste sistema era a capacidade e velocidade do meio de ligação à Internet que cada utilizador dispõe. Enquanto que do lado emissor a rede da Universidade de Aveiro possui uma ligação de alto débito, a diferenciação fez-se do lado receptor. Desta forma, e mais uma vez recorrendo a duas formas já bastante disseminadas pelos utilizadores, testámos a ligação via modem analógico de 56 Kbps e uma ligação via ADSL a 512/ 128 Kbps. Decidimos também utilizar ligações do mesmo operador dado não ser nossa intenção avaliar o serviço dos operadores de Internet mas sim a influência que a velocidade de transmissão tem para o vídeo ao vivo na Internet.

No que diz respeito aos programas de leitura de vídeo, para além dos leitores Windows Media Player [5] e Real Player [6], optámos por utilizar um leitor neutro aos dois sistemas base para avaliar as diferenças entre cada um dos produtos de codificação. A escolha recaiu no MPlayer [7], um leitor desenvolvido em Linux e adaptado a sistemas Windows, capaz de reconhecer os formatos de codificação de cada um dos codificadores utilizados. Este leitor possui também ferramentas muito úteis para este tipo de trabalho, principalmente a capacidade de guardar no PC receptor o vídeo recebido desmontado em frames.

Outro parâmetro que pretendíamos analisar era a influência da capacidade do PC onde residem os leitores de vídeo. Foram utilizadas duas máquinas com as seguintes características principais:

- a) PC portátil com processador Pentium III a 690 Mhz e memória RAM de 128 MB, que designaremos por PC A;
- b) PC portátil com processador Pentium IV a 2,8 Ghz e 224 MB de memória RAM, que designaremos por PC B;

Em ambos o sistema operativo instalado era o Windows XP SP1.

Dadas algumas dificuldades logísticas para utilizar uma webcam para captar o sinal a fornecer aos codificadores optámos por utilizar um vídeo já existente no disco duro do nosso servidor. Convém referir que este vídeo foi gravado após ter sido codificado em formato mpeg. Na apresentação dos resultados iremos apresentar a comparação de imagens recebidas após a transmissão com este vídeo original para nos apercebermos dos efeitos que o canal de transmissão e os codificadores tiveram nesta experiência.

V.2 – Descrição das experiências

Os parâmetros em avaliação nestas experiências eram a avaliação da qualidade do vídeo recebido no leitor em função de:

- ? Velocidade de ligação no lado do leitor
- ? Capacidades do PC
- ? Codificador
- ? Leitor de Vídeo

Estabelecemos oito conjuntos de experiências para os quais utilizámos sempre o mesmo vídeo à entrada do codificador:

Experiência	Codificador	PC	Ligação	Leitores
A	Windows Media Encoder	A	ADSL	Windows Media Player e MPlayer
B	Windows Media Encoder	A	Modem	Windows Media Player e MPlayer
C	Windows Media Encoder	B	ADSL	Windows Media Player e MPlayer
D	Windows Media Encoder	B	Modem	Windows Media Player e MPlayer
E	Real Producer Basic	A	ADSL	Real Player e MPlayer
F	Real Producer Basic	A	Modem	Real Player e MPlayer
G	Real Producer Basic	B	ADSL	Real Player e MPlayer
H	Real Producer Basic	B	Modem	Real Player e MPlayer

Tabela V-1 – Conjunto de experiências realizadas

Em cada uma destas experiências fazíamos ainda variar no codificador a taxa de bits do vídeo transmitido o que resultou em oitenta situações distintas. A avaliação do desempenho de cada uma das situações era feita por dois métodos de análise:

- ? Análise qualitativa do vídeo apresentado no leitor baseada na visualização das imagens decodificadas;
- ? Análise quantitativa através da comparação do número de *frames* enviadas pelo servidor de *streaming* com o número de *frames* do vídeo à saída do leitor.

V.3– Resultados

Passamos a apresentar os resultados obtidos para cada um dos 8 conjuntos de experiências acima indicados:

a) Windows Media Encoder, leitor no PC A ligado por ADSL

Análise quantitativa:

Identificação da experiência	Codificador				Leitor MPlayer		
	Velocidade Esperada (kbps)	Velocidade efectiva (Kbps)	Fps esperado	Fps efectivo	Fps recebido	Recebidos vs enviados	Tamanho da imagem
A1	28	26,55	12,5	12,50	12,32	98,53 %	192 x 144
A2	86	85,29	12,5	12,50	11,98	95,87 %	288 x 216
A3	121	119,18	12,5	12,50	12,50	100,00 %	384 x 288
A4	241	237,92	25,0	24,99	19,72	78,89 %	384 x 288
A5	491	488,26	25,0	24,99	19,17	76,70 %	384 x 288
A6	691	688,95	25,0	24,99	14,17	56,70 %	384 x 288

Tabela V-2 – Análise Quantitativa da experiência a)

Análise qualitativa:

Identificação da experiência	Codificador		Leitor	
	Velocidade de codificação (kbps)	Fps esperado	Windows Media Player	MPlayer
A1	28	12,5	Imagem pequena com contornos de objectos deficientes. Sem paragens.	Imagem pequena com contornos de objectos deficientes. Sem paragens.
A2	86	12,5	Imagem boa com ligeiros erros nos limites dos objectos. Notam-se os quadrados dos macroblocos. Sem paragens.	Imagem boa com ligeiros erros nos limites. Notam-se os quadrados dos macroblocos. Sem paragens.
A3	121	12,5	Imagem boa com ligeiros erros. Sem paragens	Imagem boa com ligeiros erros. Sem paragens
A4	241	25,0	Óptima imagem. Sem paragens.	Óptima imagem. Sem paragens.
A5	491	25,0	Erro – Catastrophic failure	Óptima definição em cada frame . Ligeiras paragens. Frequentes saltos.
A6	691	25,0	Erro – Catastrophic failure	Boa qualidade individual das frames. Arrastamento da imagem com paragens

Tabela V-3 – Análise Qualitativa da experiência a)

b) Windows Media Encoder, leitor no PC A ligado por Modem

Análise quantitativa:

Identificação da experiência	Codificador				Leitor Mplayer		
	Velocidade Esperada (kbps)	Velocidade efectiva (Kbps)	Fps esperado	Fps efectivo	Fps recebido	Recebidos vs enviados	Tamanho da imagem
B1	16	15,05	12,5	11,23	10,13	90,23 %	192 x 144
B2	28	26,50	12,5	12,49	12,35	98,88 %	192 x 144
B3	40	38,70	12,5	12,50	11,88	95,07 %	192 x 144
B4	86	84,18	12,5	12,50	19,72	46,27 %	288 x 216

Tabela V-4 – Análise Quantitativa da experiência b)

Análise qualitativa:

Identificação da experiência	Codificador		Leitor	
	Velocidade de codificação (kbps)	Fps esperado	Windows Media Player	MPlayer
B1	16	12,5	Imagem pequena e de qualidade deficiente. Sem paragens.	Imagem pequena e de qualidade deficiente. Sem paragens.
B2	28	12,5	Imagem pequena com contornos de objectos deficientes. Sem paragens.	Imagem pequena com contornos de objectos deficientes. Sem paragens.
B3	40	12,5	Imagem com qualidade razoável. Algumas paragens.	Imagem com qualidade razoável. Poucas paragens.
B4	86	12,5	Catastrophic Failure	Imagem com alguma qualidade. Paragens frequentes, algumas prolongadas

Tabela V-5 – Análise Qualitativa da experiência b)

c) Windows Media Encoder, leitor no PC B ligado por ADSL

Análise quantitativa:

Identificação da experiência	Codificador				Leitor MPlayer		
	Velocidade Esperada (kbps)	Velocidade efectiva (Kbps)	Fps esperado	Fps efectivo	Fps recebido	Recebidos vs enviados	Tamanho da imagem
C1	28	26,51	12,5	12,50	12,47	99,76 %	192 x 144
C2	86	84,00	12,5	12,50	12,50	100,00 %	288 x 216
C3	121	119,57	12,5	12,50	12,50	100,00 %	384 x 288
C4	241	237,47	25,0	24,99	24,98	99,96 %	384 x 288
C5	491	488,51	25,0	24,99	20,11	80,47 %	384 x 288
C6	691	688,48	25,0	24,99	14,67	58,70 %	384 x 288

Tabela V-6 – Análise Quantitativa da experiência c)

Análise qualitativa:

Identificação da experiência	Codificador		Leitor	
	Velocidade de codificação (kbps)	Fps esperado	Windows Media Player	MPlayer
C1	28	12,5	a)	Imagem pequena de qualidade reduzida. Sem paragens.
C2	86	12,5	a)	Imagem de qualidade razoável. Sem paragens.
C3	121	12,5	a)	Imagem com alguma qualidade, embora com erros nos limites dos objectos. Sem paragens.
C4	241	25,0	a)	Imagem razoável / boa qualidade. Sem paragens
C5	491	25,0	a)	Imagem de óptima qualidade. Notam-se alguns saltos.
C6	691	25,0	a)	Atraso na apresentação. Arrastamento das imagens. Frames individuais de óptima qualidade.

a) Não foi possível conectar via Windows Media Player com o PC B.

Tabela V-7 – Análise Qualitativa da experiência c)

d) Windows Media Encoder, leitor no PC B ligado por Modem

Análise quantitativa:

Identificação da experiência	Codificador				Leitor Mplayer		
	Velocidade Esperada (kbps)	Velocidade efectiva (Kbps)	Fps esperado	Fps efectivo	Fps recebido	Recebidos vs enviados	Tamanho da imagem
D1	16	15,03	12,5	11,16	11,10	99,46 %	192 x 144
D2	28	26,65	12,5	12,50	12,47	99,73 %	192 x 144
D3	40	38,45	12,5	12,50	11,56	92,50 %	192 x 144
D4	86	84,29	12,5	12,49	5,83	46,68 %	288 x 216

Tabela V-8 – Análise Quantitativa da experiência d)

Análise qualitativa:

Identificação da experiência	Codificador		Leitor	
	Velocidade de codificação (kbps)	Fps esperado	Windows Media Player	MPlayer
D1	16	12,5	a)	Imagem pequena com atraso na apresentação. Fraca qualidade. Sem paragens.
D2	28	12,5	a)	Imagem pequena e de qualidade reduzida. Sem paragens.
D3	40	12,5	a)	Imagem razoável com erros nos contornos dos objectos. Sem paragens.
D4	86	12,5	a)	Imagem de qualidade boa. Paragens frequentes de alguma duração.

a) Não foi possível conectar via Windows Media Player com o PC B

Tabela V-9 – Análise Qualitativa da experiência d)

e) Real Producer, leitor no PC A ligado por ADSL

Análise quantitativa:

Identificação da experiência	Codificador				Leitor MPlayer		
	Velocidade Esperada (kbps)	Velocidade efectiva (Kbps)	Fps esperado	Fps efectivo	Fps recebido	Recebidos vs enviados	Tamanho da imagem
E1	20	19	15	5,41	5,30	98,19 %	256 x 208
E2	50	49	15	8,28	8,175	98,73 %	352 x 288
E3	100	99	15	12,51	12,40	99,09 %	352 x 288
E4	225	224	30	24,99	22,50	90,06 %	352 x 288
E5	450	448	30	24,99	14,86	59,50 %	352 x 288
E6	700	697	30	24,99	11,26	45,08 %	352 x 288

Tabela V-10 – Análise Quantitativa da experiência e)

Análise qualitativa:

Identificação da experiência	Codificador		Leitor	
	Velocidade de codificação (kbps)	Fps esperado	Real Player	MPlayer
E1	20	15	Atraso na apresentação. Qualidade deficiente. Imagem com objectos muito mal definidos. Sem paragens.	Atraso na apresentação. Qualidade deficiente. Imagem com objectos muito mal definidos.
E2	50	15	Imagem com contornos mal definidos. Sem paragens.	Imagem com contornos mal definidos. Sem paragens.
E3	100	15	Imagem de boa qualidade. Ligeiros erros nos limites dos objectos. Sem paragens.	Imagem de boa qualidade. Ligeiros erros nos limites dos objectos. Sem paragens.
E4	225	30	Imagem com boa definição. Algumas paragens.	Imagem com boa definição. Sem paragens ou falhas
E5	450	30	Imagem com alguma qualidade. Paragens prolongadas frequentes.	Imagem com alguma qualidade mas com sobreposição de imagens em algumas situações de transição. Paragens se houver movimentos
E6	700	30	Imagem com deficiente qualidade. Paragens prolongadas frequentes. A apresentação ficou estática a meio da transmissão.	Má qualidade. Sobreposição frequente de imagens na mesma frame. Com movimentos é impossível de visualizar. Paragens prolongadas quando há movimento.

Tabela V-11 – Análise Qualitativa da experiência e)

f) Real Producer, leitor no PC A ligado por Modem

Análise quantitativa:

Identificação da experiência	Codificador				Leitor Mplayer		
	Velocidade Esperada (kbps)	Velocidade efectiva (Kbps)	Fps esperado	Fps efectivo	Fps recebido	Recebidos vs enviados	Tamanho da imagem
F1	20	19	15	5,40	5,21	96,46 %	256 x 208
F2	34	33	15	5,41	5,24	97,00 %	352 x 288
F3	50	49	15	8,28	3,71	44,86 %	352 x 288
F4	100	100	15	12,49	3,01	24,08 %	352 x 288

Tabela V-12 – Análise Quantitativa da experiência f)

Análise qualitativa:

Identificação da experiência	Codificador		Leitor	
	Velocidade de codificação (kbps)	Fps esperado	Real Player	MPlayer
F1	20	15	Atraso na apresentação. Imagem com deficiente qualidade e arrastamento.	Atraso na apresentação. Imagem com deficiente qualidade e arrastamento.
F2	34	15	Menor atraso na apresentação. Sobreposição de imagens em algumas frames. Poucas paragens.	Menor atraso na apresentação. Sobreposição de imagens em algumas frames. Sem paragens.
F3	50	15	Imagem de qualidade razoável. Melhor definição. Notam-se os macroblocos. Algumas paragens.	Ligeiro atraso na apresentação. Imagem de qualidade razoável. Melhor definição. Notam-se os macroblocos. Poucas paragens.
F4	100	15	Sobreposição de imagens em algumas frames. Imagem parada com frequência. Sem movimento a imagem é boa.	Sobreposição de imagens em algumas frames. Imagem parada com frequência com corte transversal a meio. Arrastamento frequente que impossibilita a visualização.

Tabela V-13 – Análise Qualitativa da experiência f)

g) Real Producer, leitor no PC B ligado por ADSL

Análise quantitativa:

Identificação da experiência	Codificador				Leitor MPlayer		
	Velocidade Esperada (kbps)	Velocidade efectiva (Kbps)	Fps esperado	Fps efectivo	Fps recebido	Recebidos vs enviados	Tamanho da imagem
G1	20	19	15	5,42	5,23	96,47 %	256 x 208
G2	50	49	15	8,28	8,26	99,81 %	352 x 288
G3	100	100	15	12,51	12,51	100,00 %	352 x 288
G4	225	225	30	24,96	24,95	99,96 %	352 x 288
G5	450	451	30	24,99	20,85	83,47 %	352 x 288
G6	700	698	30	24,99	9,79	39,18 %	352 x 288

Tabela V-14 – Análise Quantitativa da experiência g)

Análise qualitativa:

Identificação da experiência	Codificador		Leitor	
	Velocidade de codificação (kbps)	Fps esperado	Real Player	MPlayer
G1	20	15	Atraso na apresentação. Qualidade deficiente. Frequentes saltos.	Atraso na apresentação. Qualidade deficiente. Frequentes paragens e saltos.
G2	50	15	Qualidade fraca. Arrastamento e lentidão na apresentação. Nota-se a existência de poucas frames	Qualidade fraca. Arrastamento e lentidão na apresentação. Nota-se a existência de poucas frames.
G3	100	15	Imagem de qualidade razoável. Imagem sem paragens.	Imagem de qualidade razoável. Imagem sem paragens.
G4	225	30	Imagem razoável que é boa se houver pouco movimento. Vídeo sem paragens.	Imagem razoável que é boa se houver pouco movimento. Notam-se os macroblocos. Imagem sem paragens.
G5	450	30	Imagem razoável que é boa se houver pouco movimento. Notam-se os macroblocos. Paragens prolongadas.	Imagem razoável que é boa se houver pouco movimento. Notam-se os macroblocos. Imagem com poucos saltos.
G6	700	30	Frequentemente sobreposições de imagens na mesma frame. Paragens prolongadas. Ficou parado por volta do meio da apresentação.	Imagem aos saltos. Frequentemente sobreposições de imagens na mesma frame. Visualização difícil.

Tabela V-15 – Análise Qualitativa da experiência g)

h) Real Producer, leitor no PC B ligado por Modem

Análise quantitativa:

Identificação da experiência	Codificador				Leitor Mplayer		
	Velocidade Esperada (kbps)	Velocidade efectiva (Kbps)	Fps esperado	Fps efectivo	Fps recebido	Recebidos vs enviados	Tamanho da imagem
H1	20	19	15	5,42	5,27	97,18 %	256 x 208
H2	34	34	15	5,62	5,40	96,09 %	352 x 288
H3	50	49	15	8,28	6,96	84,04 %	352 x 288
H4	100	99	15	12,43	4,76	38,35 %	352 x 288

Tabela V-16 – Análise Quantitativa da experiência h)

Análise qualitativa:

Identificação da experiência	Codificador		Leitor	
	Velocidade de codificação (kbps)	Fps esperado	Real Player	MPlayer
H1	20	15	Algum atraso. Qualidade deficiente. Razoável se houver pouco movimento. Poucas paragens.	Elevado atraso. Qualidade deficiente. Razoável se houver pouco movimento. Poucas paragens.
H2	34	15	Qualidade razoável. Notam-se os macroblocos. Alguns saltos.	Menor atraso. Qualidade razoável. Notam-se os macroblocos. Algumas paragens.
H3	50	15	Imagem com paragens como se só houvesse 1 FPS.	Qualidade razoável. Imagem com sobreposições na mesma frame. Paragens reduzidas.
H4	100	15	Imagem com paragens como se só houvesse 1 FPS.	Imagem com muitas falhas. Impossível visualizar movimentos. Mesmo parada, a imagem fica arrastada verticalmente.

Tabela V-17 – Análise Qualitativa da experiência h)

V.3 – Análise dos Resultados

Influência da velocidade de acesso

Os resultados apresentados demonstram claramente que a velocidade de acesso à Internet do cliente da nossa transmissão é o principal factor de diferenciação da qualidade final da sessão de vídeo na Internet.

Com a ligação por modem analógico o débito de codificação máximo para viabilizar a transmissão é inferior a 50 Kbps. Com esta velocidade a qualidade das imagens é ainda bastante reduzida. Vejamos dois exemplos:



Fig. V-2 – Exemplo de uma frame do vídeo original á entrada do codificador.



Fig. V-3 – Exemplo de frame codificada pelo Windows Media Encoder a 40 Kbps e recebida via modem



Fig. V-4 – Exemplo de frame codificada pelo Real Producer Basic a 50 Kbps e recebida via modem

Com o Windows Media Encoder a imagem é muito pequena e não são perfeitamente visíveis os contornos dos objectos tal como se pode observar no quadro pendurado na parede atrás da figura humana presente. O Real Producer produz uma imagem de tamanho superior, quase igual à original, mas a qualidade após a transmissão deixa muito a desejar.

Para além da própria qualidade das *frames* recebidas, também o decorrer da apresentação fica muito afectado para débitos mais elevados. As paragens tornam-se frequentes e os saltos uma constante. Em algumas *frames* notam-se perfeitamente os efeitos das distorções causadas pelo estreito canal de transmissão:



Fig. V-5 – Exemplo de frame distorcida. Codificação pelo Windows Media Encoder a 40 Kbps e recebida via modem



Fig. V-6 – Exemplo de frame distorcida. Codificação pelo Real Producer Basic a 50 Kbps e recebida via modem

Com ligações ADSL a situação é bastante diferente. O débito de codificação suportado pode subir até aos 500 Kbps e o tamanho da imagem descodificada fica semelhante ao da original:



Fig. V-7 – Exemplo de frame do vídeo original á entrada do codificador.



Fig. V-8 – Exemplo de frame codificada pelo Windows Media Encoder a 491 Kbps e recebida via ADSL



Fig. V-9 – Exemplo de frame codificada pelo Real Producer Basic a 450 Kbps e recebida via ADSL

A qualidade obtida à saída do decodificador é muito pouco diferente da imagem original para qualquer dos dois codificadores utilizados.

O mesmo tipo de distorção que verificámos na ligação por modem também se verifica na ligação por ADSL. Contudo, apenas quando o débito ultrapassa os 500 Kbps como podemos verificar nas seguintes imagens:



Fig. V-10 – Exemplo de frame com sobreposição de imagens. Codificação pelo Windows Media Encoder a 691 Kbps e recebida via ADSL



Fig. V-11 – Exemplo de frame distorcida. Codificação pelo Real Producer Basic a 700 Kbps e recebida via ADSL

Quando há grandes variações ou movimentos nas imagens estas distorções são bastante mais frequentes.

Concluimos que apenas um acesso à Internet de Banda Larga, como é o caso do ADSL, pode garantir uma transmissão de vídeo em tempo real com qualidade de imagem e sem perturbações de transmissão.

Influência da capacidade do PC

Se compararmos o desempenho da transmissão do vídeo em tempo real nos dois computadores utilizados, encontramos menos diferenças do que as esperadas antes da realização da experiência.

As taxas de *frames* perdidas são praticamente iguais apenas com uma diferença importante: nos débitos de codificação limites para cada meio de recepção (500 Kbps no ADSL e 50 Kbps para o modem analógico) notam-se diferenças substanciais. Vejamos os seguintes gráficos que relacionam a taxa de *frames* perdidas para o mesmo tipo de ligação em função do débito de codificação para cada um dos pares PC/codificador utilizado:

Ligação por ADSL

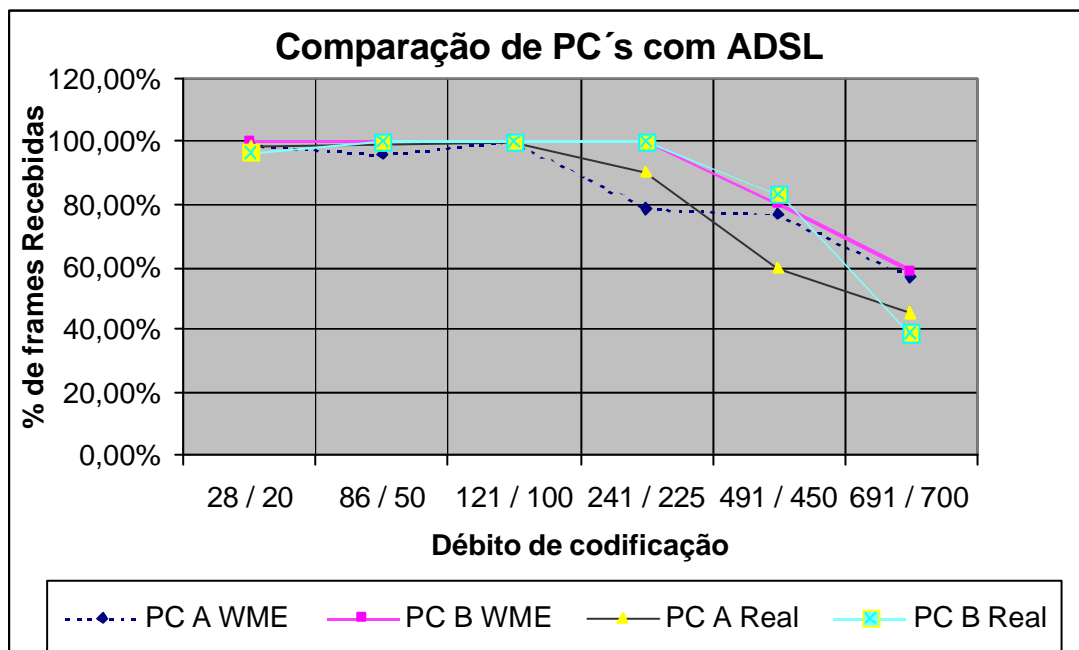


Fig. V-12 – Comparação da percentagem de frames recebidas por cada PC ligado por ADSL

Tanto no caso do Windows Media Encoder como no Real Producer Basic, apenas temos diferenças nos débitos entre 200 Kbps e 500 Kbps, o que nos leva a concluir que as superiores características do PC B são importantes nas situações limite em que há necessidade de processar maiores volumes de informação e a memória e velocidade do processador podem fazer a diferença.

Ligação por modem analógico

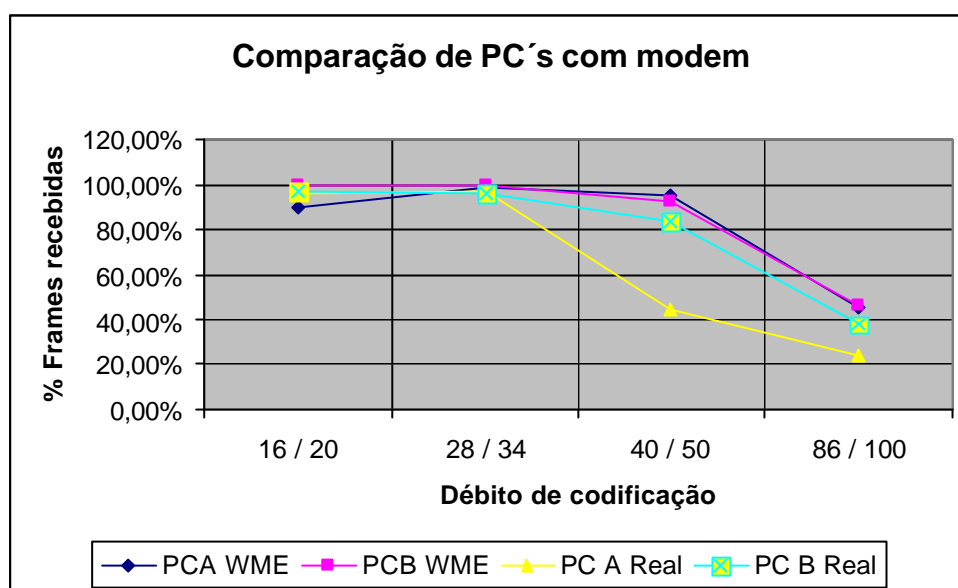


Fig. V-13 – Comparação da percentagem de frames recebidas por cada PC ligado por modem

Na recepção via modem analógico, devido ao menor tamanho da imagem e, consequentemente, do débito binário inferior que é gerado, ambos os PC's foram capazes de ter um desempenho muito aproximado. Com o Windows Media Encoder nem sequer há diferenças visíveis. No caso do Producer Basic constatámos que a 50 Kbps se verifica o mesmo fenómeno já explicado no ADSL: na velocidade limite, o decodificador Mplayer no PC B conseguiu decodificar e apresentar um número de *frames* bastante superior ao do PC A.

Concluindo, há vantagens em utilizar um PC com desempenho superior nos casos de vídeos codificados a débitos próximos do limite da nossa ligação. Essa vantagem é mais evidente nas ligações de alto débito.

Comparação entre codificadores

A nossa intenção de fazer uma comparação dos codificadores para uma transmissão de vídeo em tempo real ficou muito condicionada devido às premissas que referimos no início deste capítulo.

Ao utilizarmos as versões gratuitas dos codificadores limitámos as opções dos mesmos o que não permitiu estabelecer cenários idênticos para uma comparação directa.

No Windows Media Encoder a selecção de um dado débito de codificação origina um vídeo codificado a uma taxa de imagens por segundo que é constante, sendo a dimensão da imagem ajustada para esse objectivo. No Real Producer Basic, a mesma selecção origina um vídeo cuja taxa de imagens por segundo pode variar para se manter ao longo da codificação o tamanho e a qualidade das imagens individuais. Desta forma, para um

mesmo débito de sinal codificado, temos duas taxas de imagens por segundo completamente distintas. Por exemplo, na experiência A1, obtivemos 12,5 fps (*frames por segundo*) no Windows e apenas 5,4 fps no Real. No entanto, o tamanho da imagem neste último era bastante superior ao tamanho da imagem obtida pelo codificador da Microsoft. Apenas com a versão Real Producer Plus [8] se poderia definir um cenário que se equivalesse em absoluto com um dos cenários do Windows Media Encoder. Contudo, alguns dos cenários estudados têm parâmetros semelhantes e podem ser usados para uma comparação.

No caso da transmissão por ADSL apresentamos em seguida o comparativo dessas experiências:

Para o PC A:

Experiências comparadas	PC Utilizado	Codificador			Observações com o MPlayer	
		Velocidade de codificação (kbps)	Fps obtido	Imagem	Real Producer	Windows Media Encoder
A3 / E3	A	121 / 100	12,50 / 12,51	384x228 / 352x228	Imagem de boa qualidade. Ligeiros erros nos limites dos objectos. Sem paragens.	Imagem boa com ligeiros erros. Sem paragens
A4 / E4	A	241 / 225	24,99 / 24,99	384x228 / 352x228	Imagem com boa definição. Sem paragens ou falhas	Óptima imagem. Sem paragens.
A5 / E5	A	491 / 450	24,99 / 24,99	384x228 / 352x228	Imagem com alguma qualidade mas com sobreposição de imagens em algumas situações de transição. Paragens se houver movimentos	Óptima definição em cada frame . Ligeiras paragens. Frequentes saltos.
A6 / E6	A	691 / 700	24,99 / 24,99	384x228 / 352x228	Má qualidade. Sobreposição frequente de imagens na mesma frame. Com movimentos é impossível de visualizar. Paragens prolongadas quando há movimento.	Boa qualidade individual das frames. Arrastamento da imagem com paragens.

Tabela V-18 – Comparação entre codificadores no PC A e ligação por ADSL

Registamos diferenças mais visíveis nas velocidades de codificação mais elevadas. Verificamos que, para codificações próximas de 500 Kbps, o limite da ligação do cliente, a qualidade das imagens é melhor na transmissão gerada pelo sistema Windows, especialmente quando há transições e movimentos. Vejamos um exemplo:



Fig. V-14 – Exemplo de frame do vídeo original á entrada do codificador.



Fig. V-15 – Exemplo de imagem obtida por codificação a 491 Kbps pelo Windows Media Encoder



Fig. V-16 – Exemplo de imagem obtida por codificação a 450 Kbps pelo Real Producer Basic

Quando a codificação ultrapassa a velocidade de ligação, a percentagem de perda de pacotes sobe bastante e o decorrer do vídeo é muito afectado em ambos os casos. Como o número de imagens processadas pelo leitor foi superior para o Windows Media Encoder, este comportou-se melhor, com menos paragens.

E no caso do PC B:

Experiências comparadas	PC Utilizado	Codificador			Observações com o MPlayer	
		Velocidade de codificação (Kbps)	Fps obtido	Imagem	Real Producer	Windows Media Encoder
C3 / G3	B	121 / 100	12,50 / 12,51	384x228 / 352x228	Imagem de qualidade razoável. Apresentação sem paragens.	Imagem com alguma qualidade, embora com erros nos limites dos objectos. Sem paragens.
C4 / G4	B	241 / 225	24,99 / 24,96	384x228 / 352x228	Imagem razoável que é boa se houver pouco movimento. Vídeo sem paragens.	Imagem razoável / boa qualidade. Sem paragens.
C5 / G5	B	491 / 450	24,99 / 24,99	384x228 / 352x228	Imagem razoável que é boa se houver pouco movimento. Notam-se os macroblocos de codificação. Vídeo com poucos saltos.	Imagem de óptima qualidade. Notam-se alguns saltos.
C6 / G6	B	691 / 700	24,99 / 24,99	384x228 / 352x228	Imagem aos saltos. Frequentes sobreposições de imagens na mesma frame. Visualização difícil.	Atraso na apresentação. Arrastamento das imagens. Frames individuais de óptima qualidade.

Tabela V-19 – Comparação entre codificadores no PC B e ligação por ADSL

As principais diferenças notam-se na velocidade próxima aos 500 Kbps – experiências C5 e G5 – onde o Windows Media Encoder teve melhor comportamento ao nível de qualidade de imagem. Como exemplo apresentamos as imagens V-17, V-18 e V-19.

O decorrer do vídeo tem um comportamento diferente do observado com o PC de menores recursos. Embora na taxa de perda de *frames* na transmissão as diferenças sejam mínimas (ver tabelas V-6 e V-14), o leitor conseguiu receber e tratar mais *frames* na codificação feita pelo Real Producer o que fez com que a apresentação do vídeo não tenha tantas interrupções e saltos como no caso do Windows Media Encoder.



Fig. V-17 – Exemplo de uma frame do vídeo original á entrada do codificador.



Fig. V-18 – Exemplo de imagem obtida por codificação a 491 Kbps pelo Windows Media Encoder



Fig. V-19 – Exemplo de imagem obtida por codificação a 450 Kbps pelo Real Producer Basic

Na ligação por modem apenas a experiência com débito de codificação próximo a 100 Kbps tinha parâmetros semelhantes. Como a esta velocidade a perda de *frames* na transmissão foi elevada, não é viável fazermos qualquer avaliação comparativa da qualidade obtida.

Em resumo, podemos concluir que não há diferenças significativas entre os dois codificadores, excepto nas codificações com débitos próximos do limite da velocidade de ligação do leitor, onde o Windows obtém melhor qualidade de imagens, embora tenhamos resultados diferentes para cada PC no número de *frames* por segundo que o leitor processou para cada um.

Capítulo VI – Conclusões

A transmissão de vídeo em tempo real na Internet de uma maneira fácil, funcional e acessível à maioria dos utilizadores da Web é um desígnio perseguido desde há algum tempo. A forma como a Internet se desenvolveu, possibilitou o seu aumento exponencial, por permitir diversas formas de acesso, mas trouxe, como consequência, uma heterogeneidade que obriga os implementadores da tecnologia de transmissão de vídeo a terem de criar soluções diferentes para os diferentes clientes. Também a natureza instável ao nível do comportamento, muito dependente do volume de solicitações, faz com que, em momento algum, se possa saber antecipadamente qual será o seu desempenho. Por isso, o desenvolvimento de soluções de transmissão de vídeo em tempo real na Internet defronta várias dificuldades.

Há já alguns anos que várias entidades e investigadores têm estudado e desenvolvido soluções e ferramentas específicas para este fim. Entre elas, a área que, provavelmente, mais evoluiu foi a codificação de vídeo. Desde que o ITU especificou as primeiras normas de videoconferência, um conjunto de ferramentas revelaram-se imprescindíveis a qualquer sistema deste género, tal como vimos nos Capítulos anteriores.

O historial das mais importantes normas de codificação de vídeo poderia ser sistematizado com os seguintes factos:

- ✍ A norma H.263 foi a primeira norma a dar a possibilidade, aos utilizadores com velocidades inferiores a 64 Kbps, de acederem às técnicas de transmissão de vídeo;
- ✍ A norma H.263+, posteriormente melhorada pela versão H.263++, introduziu o conceito de escalabilidade, fundamental para responder ao comportamento instável da rede;
- ✍ As normas MPEG têm tentado alcançar o objectivo de proporcionar transmissão de vídeo na Internet a todos os consumidores. A norma MPEG-4 tem mesmo o objectivo de estender o desempenho dos codecs antecessores às baixas taxas de transmissão;
- ✍ A norma MPEG-4 introduziu, entre outras inovações, o tratamento diferenciado de uma imagem por objectos;
- ✍ O MPEG e o ITU juntaram-se para criar uma norma que resulta da conjugação das evoluções do H.263 com o MPEG-4 e que resultou no Advanced Video Coding também conhecido por H.26L, H.264 ou MPEG-4 parte 10;
- ✍ O AVC atingiu desempenhos superiores a todos os seus antecessores mas exige grande capacidade de processamento às máquinas onde residem os codificadores e decodificadores;
- ✍ Empresas comerciais produziram os seus próprios codificadores e leitores especialmente dedicados à implementação na Internet;

Se avaliarmos o desempenho das sucessivas normas pelo comprimento do *bitstream* gerado pelo codificador, pode-se constatar a rápida evolução destes sistemas quando se pretendem servir clientes com baixos débitos:

- ? O H.263 diminuiu para metade o tamanho do *bitstream* do H.261;

- ? O MPEG-4 ASP (*Advanced Simple Profile*) permite poupança de 25% em comparação com o tamanho do *bitstream* do MPEG-2 [1] e 17% em relação ao H.263 [2];
- ? O tamanho do *bitstream* do AVC pode atingir apenas metade do tamanho do *bitstream* do H.263 e reduzir em 39% o do MPEG-4 [2].

Na figura seguinte (Fig VI-1), reproduzimos um gráfico com evolução dos débitos do *bitstream* destes sistemas ao longo dos últimos anos e uma previsão para os próximos [3]:

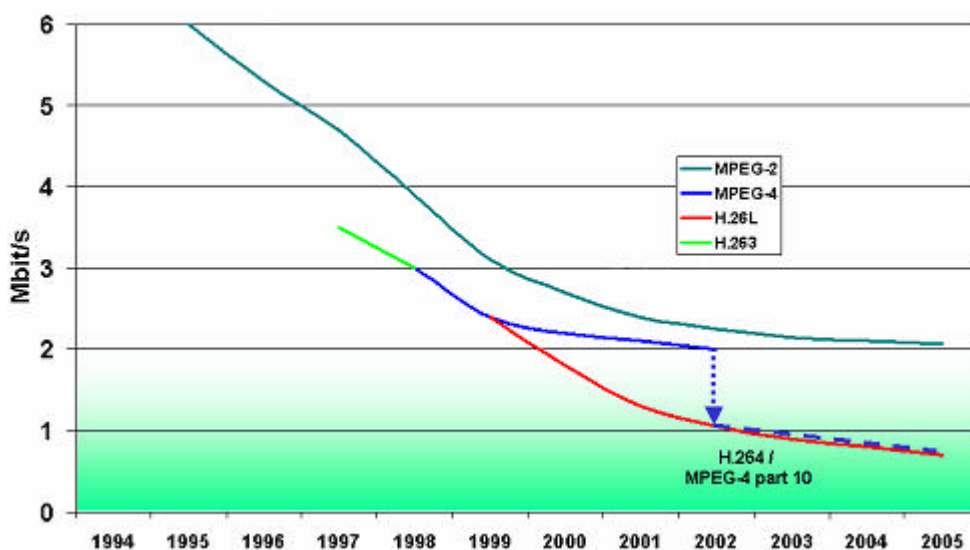


Fig VI-1 - Evolução dos débitos gerados pelas normas de codificação de vídeo [3].

Mesmo com esta progressão, é necessário que os aparelhos que suportam os leitores de streaming evoluam proporcionalmente. Por exemplo, o AVC poderá não ser ainda a ferramenta indicada para utilizadores que estão ligados à Internet com PDA's ou com telemóveis mais avançados pois as limitações dos processadores e baterias não respondem às exigências da norma.

Para além da evolução das normas de código aberto, é importante referir a contribuição das empresas que exploram comercialmente os codificadores e leitores de vídeo, tais como a Microsoft, com o Windows Media, a Apple, com a QuickTime ou a RealNetworks. Embora tenham, geralmente, desenvolvimentos proprietários, têm também de conseguir fornecer aos seus clientes produtos com melhor desempenho e utilizáveis por o maior número possível num ambiente de constante competição, o que faz com que lancem sucessivamente produtos com melhor desempenho. Todos eles alegam que conseguem melhores resultados do que as normas MPEG (ver capítulo III).

Baseado nas anteriores normas de codificação, o *Streaming* de vídeo é um conjunto de ferramentas fundamental para a transmissão de vídeo na Internet, pois fez com que não fosse necessário descarregar todo o ficheiro que constitui um vídeo para iniciar a sua apresentação. Existem mesmo desenvolvimentos específicos para *Streaming* ao nível de:

- ? Servidores;

- ? Mecanismos de distribuição de conteúdos multimédia;
- ? Adaptabilidade às condições de transmissão;
- ? Compressão e codificação de vídeo, como, por exemplo, o aperfeiçoamento das técnicas de escalabilidade como o FGS;
- ? Controlo de QoS;
- ? Sistemas implementados ao longo da rede, como o Multicast ou a replicação de conteúdo;
- ? Protocolos de rede como o RTSP.

Estes avanços permitem que já hoje se possa fazer a transmissão de vídeo em tempo real na Internet, embora sujeita a algumas condições, como um mínimo de largura de banda, reduzida taxa de erros de transmissão e atrasos ou capacidades mínimas do PC ou outro aparelho onde esteja o software de leitura de *Streaming*.

Da conjugação destes factores e do próprio desenvolvimento da rede Internet (acessos mais rápidos, maior capacidade de backbones e equipamentos activos, etc.) resultará o futuro do vídeo em tempo real. Será muito importante a massificação de acessos de Banda Larga, nomeadamente, a crescente adesão a serviços como o ADSL ou o acesso via redes de Cabo, para a difusão e aumento de consumidores de vídeo na Internet. Os dados estatísticos da Anacom (Autoridade Nacional de Comunicações) expressos na tabela VI-1, exibem a progressão de clientes destes serviços em Portugal [4]:

	1999	2000	2001	2002
Número Total de Clientes	645 146	2 110 828	3 459 640	5 165 057
Clientes de acesso <i>dial up</i>	n.d.	2 083 613	3 360 324	4 902 294
Clientes de acesso dedicado	n.d.	2 061	2 709	3 298
Clientes de acesso ADSL	n.a.	n.a.	2 886	52 005
Clientes de acesso modem por cabo	297	25 154	93 721	207 486

n.d. – não disponível; n.a. – não aplicável

Tabela VI-1 – Crescimento do número de assinantes por tipo de acesso à Internet em Portugal [4].

Verifica-se claramente o interesse crescente pelas tecnologias de Internet na população portuguesa. Enquanto os acessos dedicados, mais caros e usados quase em exclusivo por empresas e instituições, tem exibido um crescimento sustentado, os serviços ADSL e Cabo, mais acessíveis aos residenciais e pequenas empresas, estão a registar taxas de crescimento exponenciais. Para termos uma noção mais exacta deste recente crescimento, a tabela VI-2, exhibe os dados para o ano de 2003 [4]:

	2003			
	1º Trim.	2º Trim.	3º Trim.	4º Trim. ²
Número Total de Clientes	5 739 465	6 151 876	6 540 450	7 211 208
Clientes de acesso <i>dial up</i>	5 426 898	5 779 672	6 114 615	6 708 089
Clientes de acesso dedicado	3 157	3 072	3 062	3 198
Clientes de acesso ADSL	76 315	106 793	142 503	184 344
Clientes de acesso modem por cabo	233 095	262 339	280 270	315 577

Tabela VI-2 – Evolução dos acessos à Internet em Portugal no ano de 2003 [4].

Só uma base alargada de utilizadores e consumidores poderá contribuir para a rentabilidade deste tipo de transmissão. Tendo por base a experiência portuguesa, ainda algo terá de ser feito na vertente comercial para aproximar os custos destes serviços ao rendimento da esmagadora maioria da população. Mais uma vez recorrendo aos dados da Anacom (Tabela VI-3), vemos que a taxa de penetração de Banda Larga em Portugal ainda é muito reduzida apesar do crescimento acima descrito.

	2000	2001	2002
N.º de clientes Banda Larga (ADSL+Cabo)/100 Hab.	0,3%	0,9%	2,5%

	2003			
	1ºTrim	2ºTrim	3ºTrim	4ºTrim
N.º de clientes Banda Larga (ADSL+Cabo)/100 Hab.	3,0	3,5	4,1%	4,8%

Tabela VI-3 – Evolução da taxa de penetração dos acesso de Banda Larga em Portugal [4].

Segundo estes dados, no final de 2003, apenas existiam 4,8 acessos de Banda Larga por cada 100 portugueses, capazes de proporcionar experiências mais consistentes de transmissão de vídeo em tempo real. O custo mínimo deste tipo de acesso sem limites de tempo de utilização rondava até meados de 2004 os 35 € por mês, enquanto que a mensalidade para o caso de tráfego ilimitado não ficava por menos de 70 €. Este último valor representava, aproximadamente, um quinto do salário mínimo nacional. Se pensarmos que a maioria dos portugueses vivem com menos de 500 € por mês e por cabeça, apercebemo-nos de que não será fácil fazer com que haja uma grande base de clientes que suporte os serviços de *streaming*. Talvez por isso, também não há, ainda, muitos sítios de desenvolvimento específico de vídeo em tempo real para Internet. Embora por vezes se possa confundir a causa com a consequência, consideramos que se existirem clientes em número suficiente, haverá quem se interesse por lhes oferecer produtos e serviços, nem que seja devido à expectativa de retorno monetário. Muito recentemente, em finais de 2004, operadores de telecomunicações avançaram com novos serviços em ADSL cujas velocidades podem atingir os 8 Mbps no *download*. A isto acrescentaram reduções significativas de preços pois para um acesso a 2Mbps a mensalidade com tráfego limitado é de 22,5 €[5]. É um importante contributo para a ampliação do número de utilizadores. Se as velocidades anunciadas forem reais, teremos também uma importante ajuda para a transmissão de vídeo com maior qualidade.

Neste campo existe, também, alguma expectativa acerca de novos serviços que permitem o acesso à Internet de forma rápida como o acesso via rede eléctrica, já em fase de experiência comercial em Portugal, e o UMTS. São duas tecnologias que podem vir a servir quase toda a gente pois electricidade e telemóvel são quase universais.

Julgamos que a transmissão de vídeo em tempo real poderá ser a *Killer Application* que a Internet usará para se desenvolver ainda mais rapidamente, nomeadamente nos suportes em redes móveis. Se a evolução tecnológica, que tem correspondido à altura, continuar a crescer ao mesmo ritmo, e os aspectos comerciais convergirem para a realidade sócio-económica dos potenciais clientes, poderemos, num futuro a breve prazo, utilizar o nosso PC para ver notícias em directo, comunicarmos em vídeo e áudio com amigos distantes ou assistir em directo a eventos em locais longínquos com qualidade próxima da actual televisão, tudo suportado por um ambiente de Acesso Multimedia Universal (UMA) que facilitará todo o processo, mesmo para o mais leigo dos utilizadores.

Glossário

- ? **ADSL (Assymetric Digital Subscriber Line)** – Tecnologia que rentabiliza as ligações de cobre na rede telefónica tradicional de forma a poder suportar serviços de banda larga;
- ? **Backbone** – Troço principal de uma rede IP, constituída por ligações de grande capacidade;
- ? **Best Effort** – Diz-se de uma rede com recursos limitados que são distribuídos em função dos pedidos sem garantias de satisfação plena de todos.
- ? **DNS (Domain Name System)** – conjunto de regras que permite a conversão entre nomes e correspondente endereço binário na internet.
- ? **FTP (File Transfer Protocol)** – Protocolo que implementa a troca de ficheiros via Internet permitindo por exemplo o acesso reservado via uma combinação utilizador e senha.
- ? **HTTP (Hyper Text Transfer Protocol)** – Protocolo para comunicação entre computadores via Internet ou intranet que permite a visualização de páginas web garantindo a ligação entre os servidores e os clientes.
- ? **Huffman (codificação de)** – Modelo de codificação em que os códigos têm um comprimento variável e os de menor comprimento são usados para os símbolos mais frequentes para reduzir o tamanho final da informação após codificação.
- ? **IP (Internet Protocol)** – Protocolo responsável pela correcta entrega de pacotes ao longo da rede, fazendo o encaminhamento e evitando os congestionamentos.
- ? **IRC (Internet Chat)** – Conjunto de programas e canais que permitem o diálogo em tempo real via internet entre dois ou mais utilizadores da web.
- ? **ISP (Internet service Provider)** – Fornecedor do serviço de acesso à Internet;
- ? **LAN (Local Area Network)** – Rede de computadores de dimensão local;
- ? **Moore (Lei de)** – Regra empírica que diz que o crescimento da capacidade computacional dos computadores cresce ao ritmo de 2x por ano.
- ? **MPEG (Motion Picture Experts Group)** - conjunto de normas standard que define a codificação de imagens de áudio e vídeo e a sua compressão num formato digital.
- ? **POP (Point of Presence)** – Local onde um operador de telecomunicações disponibiliza um ponto de ligação à sua rede;
- ? **PC (Personal Computer)** – Computador Pessoal
- ? **PDA (Personal Digital Assistant)** – Assistente Pessoal em mini formato capaz de operar programas informáticos e simultaneamente servir de terminal móvel de telecomunicações. Vulgo Computador de Bolso;
- ? **RDIS (Rede Digital com Integração de Serviços)** – Norma para redes públicas de comunicações fixas, caracterizadas pela ligação integralmente digital da ligação do telefone do cliente até à central pública do operador. Nesta rede foram implementados vários serviços antes não permitidos nas redes analógicas (POTS) como por exemplo a identificação do chamador, números múltiplos de acesso, acesso directo a extensões, informação utilizador-a-utilizador, etc.
- ? **TCP (Transmission Control Protocol)** – Protocolo orientado à ligação responsável pela comunicação entre máquinas na internet. Encarrega-se de fragmentar a informação em pequenos pacotes que são etiquetados com informações sobre a origem, destino e tipo de informação, capazes de serem tratados pela rede.
- ? **Vídeo-on-demand** – tecnologia que permite a um cliente do serviço solicitar a um servidor remoto o fornecimento de um objecto de vídeo para visualização.
- ? **Wireless** – rede de computadores sem fios;

Bibliografia

Capítulo 1

- [1] Tanenbaum, Andrew S.: “Computer Networks”, Third Edition, Prentice Hall, 1996
- [2] Lu, Jian, “Media *Streaming* and Broadcasting on the Internet”, Apresentação para a EnjoyWeb Inc., 2000;
- [3] Conklin, Gregory J., e outros : “Video Coding for *Streaming* Media Delivery on the Internet”, IEEE Transactions on Circuits and Systems for video Technology, Vol 11, No3, Março 2001;
- [4] Wu, Dapeng, Thomas Hou, Yiwei e outros : “*Streaming* Video over the Internet: Approaches and Directions”, IEEE Transactions on Circuits and Systems for video Technology, Vol 11, No3, Março 2001;
- [5] Jornal “*Público*” de 5 de Agosto de 2003;

Capítulo 2

- [1] De Castro, F.C.C e De Castro, M.C.F.: “Codificação de canais (Capítulo V – Introdução ao sistema MPEG de Codificação de Vídeo)”, Departamento de Engenharia Eléctrica da Faculdade de Engenharia – PUCRS
- [2] Pinho, Armando: “Codificação e compressão de dados” – Apontamentos das aulas da Cadeira homónima, Dep. de Electrónica e Telecomunicações da Universidade de Aveiro
- [3] Richardson, Iain E G, “Introduction to Image and Video Coding”, www.vcodex.com, 2001- 2002;
- [4] Willrich, Prof. Roberto, “Princípios, Técnicas e Padrões de Compressão”
- [5] Sullivan, Gary : “Advanced Video Compression Standards”, Apresentação em Stanford a 15 de Fevereiro de 2001;
- [6] Guy Côté, Berna Erol, Faouzi Kossentin : “H.263+: Video Coding at Low Bit Rates”, IEEE Transactions on Circuits and Systems for Video Technology, Vol. 8, Nº 7, Novembro 1998;
- [7] ITU-T Recommendation H.263 : “Video coding for low bit rate communication” Fev. 1998
- [8] ITU-T Recommendation H.263/ Annex W : “Additional supplemental enhancement information specification”
- [9] ITU-T Recommendation H.263/ Annex U : “Enhanced reference picture selection mode”
- [10] Ribas-Corbera, Jordi, “Windows Media 9 Series – a platform to deliver compressed audio and video for Internet and broadcast applications”, EBU Technical Review – Jan 2003;
- [11] www.realnetworks.com
- [12] <http://www.microsoft.com/windows/windowsmedia/howto/articles/IntroEncoding.aspx#link2>
- [13] “Real video 10 Technical Summary”, www.realnetworks.com
- [14] KeyLabs : “RealVideo 10 Comparison Test final Report” , www.keylabs.com , 24 Dezembro de 2003;
- [15] Apple QuickTime: “QuickTime 6 Specification Sheet” : www.apple.com/QuickTime;
- [16] Li, Weiping, “Overview of Fine Granularity Scalability in MPEG-4 Video Standard”, IEEE Transactions on Circuits and Systems for video Technology, Vol 11, No3, Março 2001;
- [17] Dixon, Douglas : “*Streaming* Media: Trends and Formats”, www.manifest-tech.com, Agosto 2003;
- [18] Wiegand, Thomas e outros: “Overview of the H.264/AVC Video Coding Standard”, IEEE Transactions on Circuits and Systems for Video Technology, Julho 2003;
- [19] Saponara, Sergio e outros: “The JVT Advanced Video Coding Standard: Complexity and Performance Analysis on a Tool-by-tool Basis”, IMEC (Bélgica);
- [20] Richardson, Iain E.G., “H.264 / MPEG-4 Part 10 White Paper”, www.vcodex.com, Outubro de 2002;
- [21] Hansen, Evan : “DVD Quality Over the Internet?”, ZDNet, <http://zdnet.com.com/2100-1104-961725.html>, Outubro de 2002;
- [22] “*Streaming* Live MPEG-4 : The VBasics”, MPEG-4 *Streaming* Basics, VBrick Systems, Inc., 2003
- [23] “Report on MPEG-4 Visual fine Granularity Scalability Tools Verification Test”, ISO/IEC JTC1/SC29/WG11 Coding of Moving Pictures and Audio N4791, Maio de 2002;
- [24] “Report on the Formal Verification Tests on MPEG-4 Video Error Resilience”, ISO/IEC JTC1/SC29/WG11 Coding of Moving Pictures and Audio N2604, Dezembro de 1998;
- [25] “Report on the Formal Verification Tests on MPEG-4 Coding Efficiency for Low and Medium Bit rates”, ISO/IEC JTC1/SC29/WG11 Coding of Moving Pictures and Audio N2826, Julho de 1999;

- [26] Koenen, Rob, “MPEG-4 Overview”, ISO/IEC JTC1/SC29/WG11 Coding of Moving Pictures and Audio N4668, Março de 2002;
- [27] Grupo de Comunicações por Computador : “Fundamentos de Vídeo”, Acetatos de Sistemas Telemáticos, Universidade do Minho - Departamento de Informática
- [28] Assunção, Pedro A. Amado : “MPEG-2 Video”, Instituto de Telecomunicações, Acetatos para o Instituto Politécnico de Leiria;
- [29] Vatolin, Dmitriy e Kubasov, Denis : “VirtualDub MSU Motion Estimation Filter”, http://compression.graphicon.ru/video/motion_estimation/index_en.html

Capítulo 3

- [1] Conklin, Gregory J., e outros : “Video Coding for *Streaming* Media Delivery on the Internet”, IEEE Transactions on Circuits and Systems for video Technology, Vol 11, No3, Março 2001;
- [2] Wu, Dapeng, Thomas Hou, Yiwei e outros : “*Streaming* Video over the Internet: Approaches and Directions”, IEEE Transactions on Circuits and Systems for video Technology, Vol 11, No3, Março 2001;
- [3] Li, Weiping, “Overview of Fine Granularity Scalability in MPEG-4 Video Standard”, IEEE Transactions on Circuits and Systems for video Technology, Vol 11, No3, Março 2001;
- [4] van der Schaar, Mihaela, “A Hibrid Temporal-SNR Fine Granular Scalability for Internet Video”, IEEE Transactions on Circuits and Systems for video Technology, Vol 11, No3, Março 2001;
- [5] Wu, Feng; Li, Shipeng and Zhang, Ya -Qin : “A Framework for Efficient Progressive Fine Granularity Scalable Video Coding”, IEEE Transactions on Circuits and Systems for video Technology, Vol 11, No3, Março 2001;
- [6] “Report on MPEG-4 Visual fine Granularity Scalability Tools Verification Test”, ISO/IEC JTC1/SC29/WG11 Coding of Moving Pictures and Audio N4791, Maio de 2002;
- [7] “Report on the Formal Verification Tests on MPEG-4 Coding Efficiency for Low and Medium Bit rates”, ISO/IEC JTC1/SC29/WG11 Coding of Moving Pictures and Audio N2826, Julho de 1999;
- [8] Girod, Bernd e outros: “Advances in Channel-adaptive Video *Streaming*”, Department of Electrical Engineering, Stanford University;
- [9] Monteiro, Jânio M.E.F. e Nunes, Mário Serafim : “Distribuição de Vídeo ao Vivo em Multiponto sobre redes IP Heterogêneas”, Actas da 5ª Conferência sobre Redes de Computadores, FCCN, Setembro de 2002;
- [10] Scott, Matthew : “High Bit and Low Bit Rate Video Streamig”, Conferência Anula da CCUMC, Denton 2000;

Capítulo 4

- [1] Viana, Aline C.; Jukemura, Anibal S.e outros: “Perspectivas sobre a qualidade de Serviço nos Protocolos da Internet – Estudo de caso: Aplicações de Vídeo Sob Demanda”, Ápice Telemática, LTDA;
- [2] Zhang, Fei e outros: “*Streaming* MPEG-4 Video Over Differentiated Service Networks”, School of Electrical Engineering, University College ADFA (UNSW);
- [3] Koenen, Rob: “From MPEG-1 to MPEG-21: Creating an Interoperable Multimedia Infrastructure”, ISO/IEC JTC1/SC29/WG11 Coding of Moving Pictures and Audio N4518, Dezembro de 2001;
- [4] Torres, Luís e Delp, Edward J. : “New Trends in Image and Video Compression”
- [5] Day, Neil e Martinez, José M.: “Introduction to MPEG-7 (v4.0)”, ISO/IEC JTC1/SC29/WG11 Coding of Moving Pictures and Audio N4675, Março de 2002;
- [6] Beek, Peter van e outros: “Metadata-Driven Multimedia Access”, IEEE Signal Processing Magazine, Março de 2003;
- [7] Vetro, Anthony e outros: “Video Transcodificação Architectures and Techniques: Na Overview”, IEEE Signal Processing Magazine, Março de 2003;
- [8] Bornans, Jan e outros: “MPEG-21: The 21st Century Multimedia Framework”, IEEE Signal Processing Magazine, Março de 2003;
- [9] Pereira, Fernando e Burnett, Ian: “Universal Multimedia Experiences for Tomorrow”, IEEE Signal Processing Magazine, Março de 2003;
- [10] Saponara, Sergio e outros: “The JVT Advanced Video Coding Standard: Complexity and Performance Analysis on a Tool-by-tool Basis”, IMEC (Bélgica);
- [11] Hansen, Evan : “DVD Quality Over the Internet?”, ZDNet, <http://zdnet.com.com/2100-1104-961725.html>, Outubro de 2002;
- [12] Girod, Bernd e outros: “Advances in Channel-adaptive Video *Streaming*”, Department of Electrical Engineering, Stanford University;

- [13] Monteiro, Jânio M.E.F. e Nunes, Mário Serafim : “Distribuição de Vídeo ao Vivo em Multiponto sobre redes IP Heterogéneas”, Actas da 5ª Conferência sobre Redes de Computadores, FCCN, Setembro de 2002;
- [14] Kamaci, Nejat e Altunbasak, Tucel : “Performance Comparison of the Emerging H.264 Video Coding Standard with the Existing Standards”, Georgia Institute of Technology;

Capítulo 5

- [1] <http://www.microsoft.com/windows/windowsmedia/9series/encoder/default.aspx>
- [2] <http://www.realnetworks.com/products/producer/basic.html>
- [3] <http://www.microsoft.com/windows/windowsmedia/9series/server.aspx>
- [4] http://www.realnetworks.com/products/media_delivery.html
- [5] <http://www.microsoft.com/windows/windowsmedia/default.aspx>
- [6] <http://www.real.com>
- [7] <http://www.mplayerhq.hu/homepage/design7/news.html>
- [8] <http://www.realnetworks.com/products/producer/index.html>

Capítulo 6

- [1] Sathe, Vinay e Tayer, Marc: “MPEG-4 Video : Opportunities and Challenges Ahead” <http://www.sdtelecom.org/doc/event/ACF7A31.pdf>, Julho de 2003;
- [2] Hong, Bo : “Introduction to H264” Multimedia Communications Laboratory, University of Texas, Novembro de 2002;
- [3] Jiqiang, Song: “H.264/Advanced Video Coding – A New Standard“, <http://appsrv.cse.cuhk.edu.hk/~jqsong/H264.ppt>, Outubro de 2003;
- [4] Anacom – Autoridade Nacional de Comunicações : Relatório “[Serviço de Transmissão de Dados/ Serviço de Acesso à Internet - 4º Trimestre de 2003](http://www.anacom.pt)”, www.anacom.pt;
- [5] <http://adsl.clix.pt/produtos2mg.html>

Créditos de figuras:

- Fig. II-1a a II-1d : Vatolin, Dmitriy e Kubasov, Denis : “VirtualDub MSU Motion Estimation Filter”, http://compression.graphicon.ru/video/motion_estimation/index_en.html;
- Fig. II-2 : Grupo de Comunicações por Computador : “Fundamentos de Vídeo”, Actas de Sistemas Telemáticos, Universidade do Minho - Departamento de Informática
- Fig II-9 : Koenen, Rob, “MPEG-4 Overview”, ISO/IEC JTC1/SC29/WG11 Coding of Moving Pictures and Audio N4668, Março de 2002;
- Fig II-12 : Pinho, Armando: “Codificação e compressão de dados” – Apontamentos das aulas da Cadeira homónima, Dep. de Electrónica e Telecomunicações da Universidade de Aveiro
- Fig VI-1: Jiqiang, Song: “H.264/Advanced Video Coding – A New Standard“, <http://appsrv.cse.cuhk.edu.hk/~jqsong/H264.ppt>, Outubro de 2003;

Créditos de tabelas:

- Tabela VI-1 a VI-3 : Anacom – Autoridade Nacional de Comunicações : Relatório “[Serviço de Transmissão de Dados/ Serviço de Acesso à Internet - 4º Trimestre de 2003](http://www.anacom.pt)“, www.anacom.pt;

Índice Remissivo

- ADSL, 11, 12, 15, 55, 72, 110
- ARPA, 7
- ARPANET, 7, 9
- AVC. *Consulte* Advanced Video Coding
- caching*, 65
- Codecs de Vídeo, 53
- Codificação por VLC, 22, 26
- compensação de movimento, 23, 25, 41, 44, 60
- Compensação de Movimento, 17, 19, 20
- Controlo de taxa de transmissão, 62
- Data Striping*, 54
- DCT, 21, 22, 23, 24, 25, 27, 30, 31, 41, 59, 61, 72
- Delay-Constrained Retransmission*, 62
- Diffserv, 70
- DMIF, 35, 37
- DNS, 7, 110
- Domain Name System, 110. *Veja* DNS
- EDF, 54
- Elemento Digital**, 79
- Erliest Deadline First. *Veja* EDF
- Error Concealment*, 63
- Error-Resilient Encoding*, 62
- EuropaNET, 7
- FEC, 55, 62
- FGS, 38, 39, 58, 59, 60, 61
- Fine Granularity Scalability*. *Veja* FGS
- GPRS**, 10
- GSM**, 10, 55
- H.261, 16, 24, 25, 26, 52
- H.263, 24, 25, 26, 27, 28, 40, 52, 73
- HTTP*, 10, 36, 54, 110
- Internet Protocol. *Veja* IP
- Internet Service Provider *Veja* ISP
- IntServ, 69, 70, 71
- IP, 7, 8, 9, 12, 33, 36, 37, 41, 43, 50, 63, 65, 66, 70, 71, 110
- ISP, 8, 10, 63, 64, 66, 71, 72, 110
- mirroring*, 65, 66
- Motion Picture Experts Group, 29, 110
- MPEG1, 24, 29
- MPEG-1, 1, 23, 29, 30, 33, 37, 112
- MPEG2, 29, 30
- MPEG-2, 1, 3, 29, 30, 31, 33, 40, 41, 44, 46, 47, 73, 75, 106, 112
- MPEG-21, 73, 74, 78, 79, 80
- MPEG-4, 33, 34, 35, 36, 37, 39, 40, 41, 42, 43, 44, 45, 47, 52, 58, 69, 71, 73, 74, 75, 77, 80
- MPEG-7, 70, 73, 74, 76, 77, 78, 79
- Multimedia Content Description Interface*, 76
- Network Attached Storage, 54
- NSFNET, 7
- PFGS, 60, 61
- Predição Interframe, 17
- preroll buffer*, 53
- Progressive Fine Granularity Scalability. *Veja* PFGS
- Protocolos de sessão, 67
- Pull – splitting*, 56
- Push-splitting*, 56
- Qualidade de serviço, 54
- Quantificação, 22, 24, 44
- Quicktime, 16, 43, 44
- Rate Monotonic Scheduling, 54
- Rate Shaping, 62
- RDIS**, 11, 12, 24, 37, 55, 110
- Real Player, 16, 29, 44
- Real Time Protocol**. *Veja* RTP
- RSVP, 70
- RTCP**, 67
- RTP**, 36, 67
- SAN
 - Storage Area Networks, 54
- Service Level Agreement*, 71
- Serviços diferenciados, 70
- Serviços Integrados, 69, 70
- Simulcast*, 56
- sistema visual humano**, 16
- SMIL, 44
- Stream Synchronization Protocol, 57
- Sure Stream, 55
- TCP, 7, 8, 9, 10, 54, 67, 68, 110
- Transcoding, 69, 73, 74
- Transformada Discreta de Co-seno *Veja* DCT
- Transmission Control Protocol*. *Veja* TCP
- UDP**, 8, 9, 10, 36, 54, 67
- UMA, 73, 79
- UMTS, 10, 72
- Universal Multimedia Access. *Veja* UMA
- User Data Protocol** *Veja* UDP
- video-on-demand*, 5, 77
- Windows Media, 16, 29, 43, 45, 46, 47
- Wireless, 72, 110